MODEL-BASED ESTIMATION APPLICATIONS FOR GNSS REMOTE SENSING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Karen Qing Zi Chiang

May 2016

MODEL-BASED ESTIMATION APPLICATIONS FOR GNSS REMOTE SENSING

Karen Qing Zi Chiang, Ph.D.

Cornell University 2016

The Global Navigation Satellite System (GNSS) is a valuable tool for Earth and Atmospheric Science, as it not only provides position and attitude information for remote-sensing platforms, but its variety of radio signals may also be used to probe the atmosphere. This dissertation presents a set of projects that aim to improve several of these functions by employing model-based estimation techniques. The first algorithm addresses the issue of large uncertainties in the dynamic variations of signal observables due to refraction during Earth limb-scanning. It pre-processes raw GNSS data with a Levenberg-Marquardt batch filter, and performs signal tracking with a Square-Root Extended Kalman Filter (SREKF) in a new type of combined phase-locked/delay-locked loop. This constitutes an alternate way for deducing meteorological conditions down to few metres from the terrestrial surface using GNSS signals. The second project estimates the attitude of a spinning sounding rocket carrying sensors for studying space weather. The GPS attitude determination problem for this rocket poses two major challenges that result from equipment limitations: Frequent signal data gaps due to telemetry bandwidth restrictions and only one antenna baseline vector with which to perform full, three-axis attitude determination. The first problem is circumvented by an adaptation of the algorithm from the first project, and the second by using another Levenberg-Marquardt batch filter that contains an Euler dynamics model. The last project combines refractive ray-

tracing concepts and an SREKF, utilizing both ionosonde and GNSS signals, in order to solve for the parameters of a node-based profile of the ionosphere. In addition, a trust-region-reflective algorithm, a modern form of the Levenberg-Marquardt algorithm, is used to extract electron content information from GNSS data to parameterize ionospheric irregularities resulting from an experiment that involves controlled heating of the ionosphere.

BIOGRAPHICAL SKETCH

Karen Chiang was born in Hong Kong on January 2, 1987. In 1993, her family immigrated to Vancouver, Canada, where she spent her rambunctious years. She went on to earn her Bachelor of Science degree in Applied Physics from Columbia University in 2009, with minors in Applied Math and Economics. Due to the joint forces of the Financial Crisis and Karen's childhood dream of shooting things into space, she went straight to Cornell after graduation.

Dedicated to M. Leineweber,

and to C.R.

ACKNOWLEDGEMENTS

My utmost gratitude to the members of my graduate committee. Prof. Psiaki, who has provided excellent guidance and support over the years. Prof. Campbell for being an amazing teacher. Prof. Peck for his kindness and career advice. Prof. Hysell, who has lent his invaluable expertise to my last project.

Many thanks to my lab-mates for always being willing to discuss and collaborate: Ryan Mitch, Ryan Dougherty, Rob Miceli, and most of all, Brady O'Hanlon for also saving me many times with his CASES receiver magic. Steve Powell, our senior engineer, for his help with all the hardware and data for the MICA project. Steve Taylor, from Miami University, for being extremely accommodating with sharing HAARP data.

My wonderful friends here at Cornell, who helped make the good times fun and the bad times tolerable. My greatest thanks to my parents, Beatrix Wei and King Wai Chiang, for believing I could do anything. My sister, Fiona Chiang, in whom I can always confide. Lastly, this dissertation is dedicated to Matt Leineweber – the love of my life, my best friend, and my husband-to-be – without whom I would have starved to death before it was written.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 GNSS

The Global Navigation Satellite System (GNSS) has provided precise and flexible solutions to position, navigation, and timing (PNT) problems since the US Global Positioning System (GPS) became operational in 1994 [23]. Its multi-frequency signals have a natural application in atmospheric remote sensing, as tropospheric and ionospheric effects enter into the signal models approximately linearly, and the ionospheric effects are wavelength-dependent. These effects, which are treated purely as nuisance parameters and sources of error when using GPS purely for navigation and time keeping, provide useful observables for purposes of estimating water vapor, temperature, pressure, and electron content.

The GPS constellation now consists of 32 satellites, with 12 of them broadcasting the civil signal on a new L2C frequency (1227.60 MHz), in addition to the civil L1 signal (1575.42 MHz). A second global navigation satellite system is the Russian Federation's Global Navigation Satellite System (GLONASS), which, after a hiatus, maintains 24 satellites broadcasting over a spectrum of frequencies, centred on L1 and L2. GNSS is currently advancing to the next generation of multi-frequency constellations that include the US GPS-III block, the updated Russian GLONASS, the European Union's GALILEO, and China's Beidou/COMPASS. Regional augmentation systems, such as Japan's Quasi-Zenith Satellite System (QZSS) and India's Regional Navigation Satellite Systems (IRNSS), are also emerging. As these developments occur, there will be

greater opportunity for exploiting radio signals for remote sensing applications. The focus of the following chapters are on GPS signals, although the concepts may be easily extended to other constellations within GNSS.

1.2 The GPS Signal & Atmospheric Terms

The primary GPS frequency, at which every satellite in the network broadcasts, is L1. It is supplemented in some satellites by L2 (1227.6 MHz), the use of which has been growing exponentially. A limited number also broadcast at L5 (1176.45 MHz). GPS operates on a Code Division Multiple Access (CDMA) spread-spectrum scheme, where the 50 Hz navigation data containing orbit ephemeris are encoded on top of a satellite-dependent pseudo-random number (PRN), Fig. 1.1. Receivers employ a replica of each PRN code in order to differentiate between satellites. The Coarse/Acquisition (C/A) PRN code, for civilian use, transmits at a $1.023 \times 10^6 chips/s$ data rate, while the encrypted Precise (P(Y)) code, for U.S. military use, transmits at ten times the C/A-code chipping rate.

Each GPS L1 composite signal can therefore be expressed generally as [27]

$$y(t) = d(t)\left[A_I C(t)\cos(\omega t + \phi) + A_Q P(t)\sin(\omega t + \phi)\right] \tag{1.1}$$

where $d(t)$ represents the ± 1 navigation bit stream, A_I and A_Q are the respective in-phase and quadrature signal amplitudes, $C(t)$ and $P(t)$ are the C/A and P(Y) codes, ω is the L-band carrier frequency, and ϕ is the common phase offset.

A GPS receiver generates replicas of signal components in Eq. (1.1) and correlates them with the received signal in order to lock onto it in carrier-phase/code-delay space. Based on such tracking, the receiver can compute the

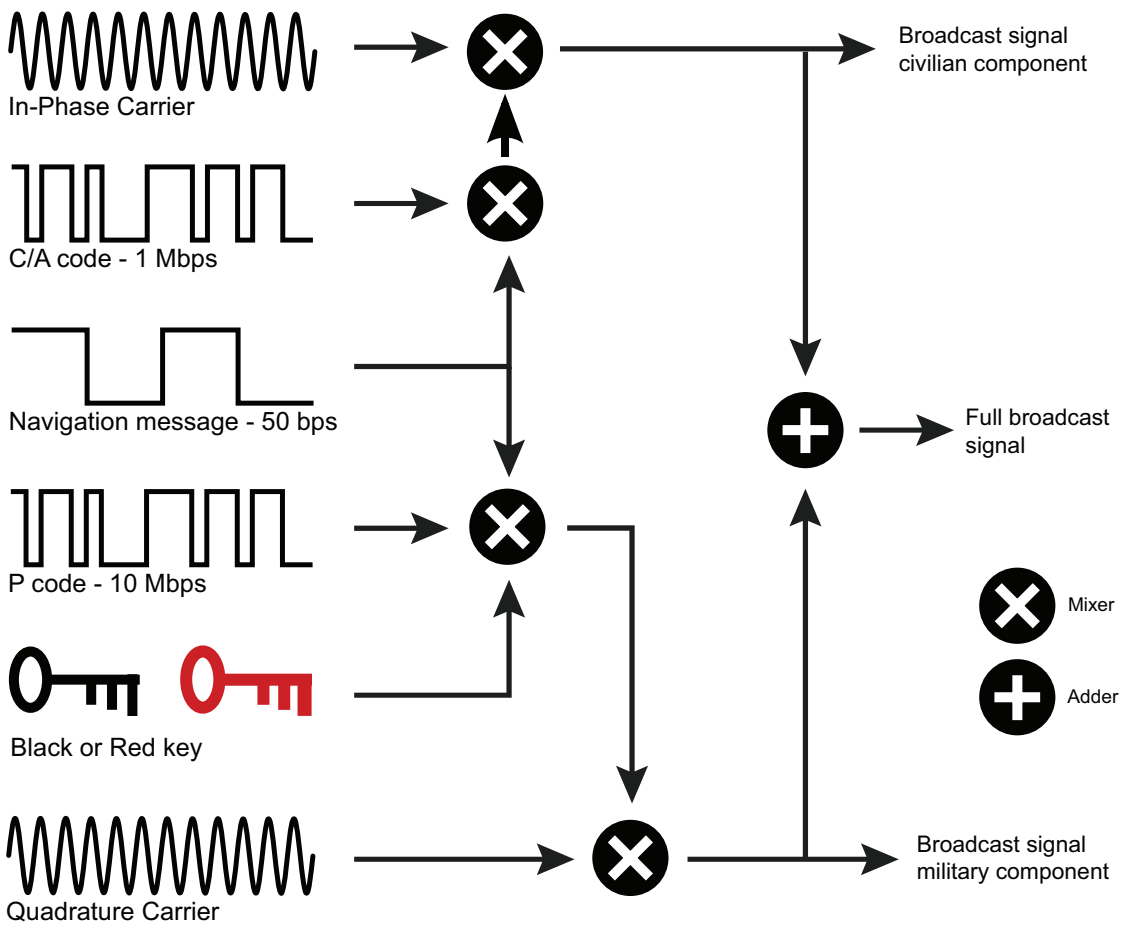

Figure 1.1: Diagram of GPS L1 signal structure. (Modified Wikipedia Commons media)

range-equivalent difference between the time of reception and time of transmission $P = c(t_r - t^s)$, called the pseudorange. This observable can be modeled as the aggregate of several factors, some of which depend on carrier frequency:

$$P(\omega) = \rho + c(\delta t_r - \delta t^s) + T + I(\omega) + M_\rho(\omega) + n_\rho(\omega) \tag{1.2}$$

where ρ is the true range, δt_r and δt^s are, respectively, the receiver and satellite clock errors, T and I are the tropospheric and ionospheric delays, M_ρ is the pseudorange multipath (reflection interference) error, and n_ρ constitutes the Gaussian sum of other delays or code errors.

The range-equivalent beat carrier phase observable, the time integral of the

carrier Doppler shift as measured and integrated by the receiver, can also be calculated and modeled as

$$\lambda\phi = \rho + c(\delta t_r - \delta t^s) + T - I(\omega) + M_\phi(\omega) + \beta\lambda + n_\phi(\omega) \tag{1.3}$$

where λ is the carrier wavelength, M_ϕ is the phase multipath error, and n_ϕ are other phase errors. β is a carrier phase bias ambiguity that enters in due to uncertainties in the initial carrier phase of the satellite transmitter electronics and in the initial carrier phase replica of the receiver signal processing electronics [38].

The tropospheric term T depends on the integrated effects of the pressure, water vapor, and temperature on the neutral atmosphere index of refraction along the line-of-sight (LOS) between the satellite and receiver. Similarly, the ionospheric term I depends on the integrated free electron density along the LOS. These refractive effects can be substantial at low elevations and can cause large errors in pseudorange and carrier phase. Fortunately, the effects can be modeled and compensated within a model-based estimation approach. Chapter 2 presents a new algorithm to track GNSS signals when these errors are hard to model *a priori* and are changing rapidly. It enables a modified receiver to track successfully, without the need for a long initial signal acquisition phase followed by a long transition phase from acquisition to steady-state tracking.

The work in Ch. 3 exploits the fact that Eq. (1.3) is especially useful for precise positioning because L-band beat carrier-phase measurements provide resolution on the order of $1/40^{th}$ of a wavelength, i.e., on the order of $0.005m$. Chapter 3 shows how carrier phase differential GPS can be harnessed for supporting sounding rocket experiments in atmospheric studies by providing spacecraft attitude.

The focus of Ch. 4 is on the larger of the effects: Ionospheric refraction, which can be quantified by using both the L1 and L2 frequencies. The interfrequency difference of pseudorange and phase varies with electron content, which is the integral of electron density along the line of sight. This leads again to the duality of I being simultaneously a degradation factor and a tool for ionospheric science. Improved mitigation of the former lends itself to more accurate characterizations for the latter. Through determining I for multiple satellites in view, doing this during controlled heating experiments that alter the ionosphere's refractive properties, and analyzing supplemental ionosonde data, Ch. 4 exploits this dual use further in order to obtain detailed pictures of the quiescent and disturbed ionosphere with GPS signals.

1.3 Dissertation Topics

The previous sections have conveyed the overarching context of applying estimation theory to utilize GNSS signals for atmospheric and space weather science. This section gives a preview, in more detail, of how the three main chapters that comprise this dissertation fit into this context, by way of a brief discussion of the goals, methods, and results of each project.

The first chapter presents an algorithm that tracks GPS observables using the output from a large number of Doppler-shift/code-phase correlators. The motivation for such an algorithm is to provide a way to use GPS signals in Earth limb-scanning. In this application, a highly dynamic, occulted GPS signal must be tracked as soon as it is seen by a receiver onboard a low Earth orbit satellite, which has little time to acquire the signal. The large uncertainties in Doppler

shift and code phase are dealt with by a sophisticated phase-locked-loop/delay-locked-loop (PLL/DLL) scheme, the main elements of which are a Levenberg-Marquardt batch filter that fits many accumulations of the received signal to a signal model, and a Kalman filter that is passed the resulting optimized signal parameters as pseudo-measurements. A limb-scan simulation models signal refraction in the atmosphere and provides the "truth" measurements, showing that with 50 Hz accumulations, the PLL has a pull-in of at minimum 122 Hz, and the DLL of 5 C/A *code chips*. The signal is tracked successfully from approximately 0.03 s after first rise.

Chapter 3 lays out an algorithm that has been developed for the three-dimensional attitude determination of a spinning sounding rocket that serves as the vehicle for space weather experiments. This algorithm overcomes challenges stemming from two primary issues: Frequent data gaps due to limited telemetry bandwidth, and only a single antenna baseline due to limited physical space. To address the first problem, GPS data are processed by a variant of the algorithm of Ch. 1. A Levenberg-Marquardt batch filter and a Rauch-Tung-Striebel (RTS) smoother fits GPS observables to the outputs of a bank of correlators. These observables are differenced between the two GPS antennas and combined with an Euler dynamics model in a separate attitude batch filter that solves a mixed real-integer optimization problem. RF data from a sounding rocket mission is used to test the algorithm of Ch. 3, yielding attitude quaternion and spin-rate estimates that have been verified with those from a magnetometer and horizon-crossing indicator to within several degrees.

The final project, delineated in Ch. 4, delves more heavily into aspects of atmospheric science. A refractive ray-tracing model of the ionospheric effects

on RF signals is used to fuse dual-frequency GPS data with ionosonde measurements in an estimation problem. A combination of filters yield the optimal parameterization for a quiescent, node-based electron density profile. GPS measurements permit estimation of the topside profile and the geospatial variation of electron content, while the ionosonde gives the bottomside profile. Dual-frequency GPS observables are also used to estimate the parameterization of a perturbed region in the ionosphere, in a separate problem that designates the quiescent profile results as nominal. Perturbations are artificially induced via controlled heating of the ionosphere above the High Frequency Active Auroral Research Program (HAARP) heater facility in Gakona, AK, and are modeled as ellipsoidal magnetic field-aligned irregularities. Experimental ionosonde and GPS data have been collected during heating campaigns and have been processed to generate quiescent and disturbed ionosphere parameterizations. The results are an average ionosonde virtual height fit error of 0.46 km and an average slant TEC error of 3.65 $TECU$ (10^{16} e^{-}/m^{2}), using six GPS satellites. The ellipsoidal perturbation model demonstrates good fits with residual norms mostly under a tenth of a $TECU$.

Within the previously stated binding theme, the three chapters also contain significant differences. The most important contrast between the first two projects is that Ch. 2 only involves an enhanced version of standard receiver signal processing for code division multiple access (CDMA) signals, whereas Ch. 3 involves model-based estimation that requires an attitude dynamics model for a spinning rocket, which is nonstandard in RF signal processing. Chapter 4 is quite unlike Chs. 2 and 3, in that the details of signal processing are less dominant, and it is assumed that the individual receivers are robust enough for its purpose. Instead, it is stepping back to examine a space weather system, esti-

mating its parameters using refraction physics models, rather than estimating simple signal parameters, as in Ch. 2, or rocket attitude and spin state parameters, as in Ch. 3.

CHAPTER 2

KALMAN FILTER TRACKING OF LIMB SCAN SIGNALS USING A BANK OF CORRELATORS

2.1 Abstract

A combined PLL/DLL algorithm is developed for tracking GNSS carrier phase and code phase using the output from a large number of correlators. This approach has advantages for limb-scanning applications, in which useful meteorological information, available only at the initial rising time of a GPS satellite, is desired. The technique uses a bank of correlators to span wide ranges of uncertainty in code phase and carrier Doppler shift, thereby avoiding the need for a separate acquisition and the associated loss of an initial span of data. A fusion of optimal estimation methods processes the output from these correlators. A batch optimization of a signal model's fit at a point in time to many accumulations from the correlator bank provides a Kalman filter with "measurements" of the most likely signal parameters, and the Kalman filter utilizes a signal dynamics model to provide estimates that drive the PLL and DLL. The effectiveness of this algorithm is demonstrated by using a truth-model simulation of a limb scan.

2.2 Introduction

GNSS receivers must achieve and maintain lock on carrier Doppler shift and the pseudo-random number (PRN) code phase in order to properly track a signal and ascertain navigation observables. A standard GNSS receiver accomplishes

this with two separate, consecutive operations: acquisition and tracking [27]. Acquisition searches for initial estimates of Doppler shift and PRN code phase. The tracking algorithm then uses these estimates to initiate a delay-locked loop (DLL) and either a frequency-locked loop (FLL) or a phase-locked loop (PLL) in order to, respectively, keep the replicas of the code and carrier signal aligned with those of the received signal.

The objective of this chapter is to create a joint PLL/DLL algorithm that functions normally even with large tracking errors and that does not require the usual transition from initial acquisition to tracking. The primary motivation for such an algorithm is to equip a low-Earth orbit (LEO) satellite, carrying a GPS receiver, with the means to capture data from a rising-GPS-satellite limb scan without any loss of data during the time it would take to carry out a standard acquisition.

To realize this goal of robustness and speed, the new tracking algorithm utilizes a bank of correlators to encompass uncertainties in carrier Doppler shift and code phase, forming rectangular regions within Doppler-shift/code-phase space [11]. Figure 2.1 illustrates this concept. Note that ω_D indicates the Doppler shift axis, and t_s indicates the PRN code delay axis. The vertical separation of the correlator banks for separate samples is not shown to scale, in order to make the figure clear. The exaggerated separation implies large changes of Doppler shift during each accumulation interval, which does not conform to the cases considered here.

In the space of these two signal properties, with time as a parameter, the estimated trajectory begins at the first estimates of Doppler shift and code phase. The uncertainty at this time is relatively large, but the bank of correlators spans

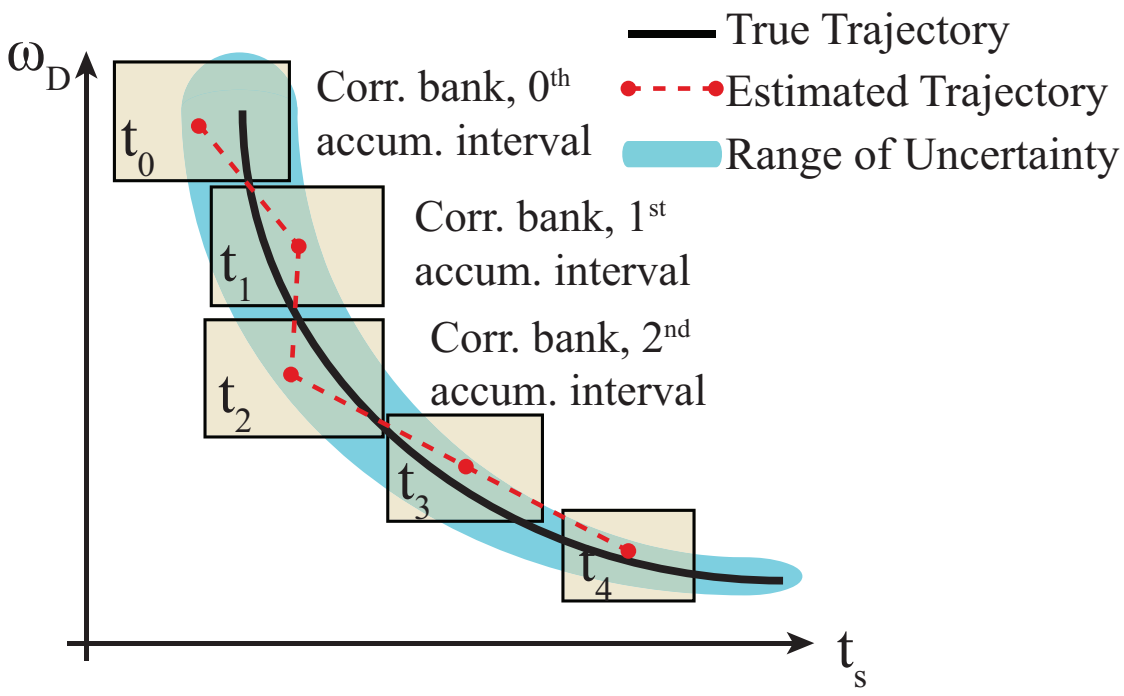

Figure 2.1: Using correlator banks to cover the true trajectory and range of uncertainty in Doppler-shift/code-phase space, an illustration.

a range that contains the true point. As time progresses, the level of uncertainty may change. In the case depicted, the uncertainty level shrinks, and therefore the size of correlator bank will also decrease.

This chapter's method seeks to continuously track time-varying carrier phase, carrier Doppler shift, carrier amplitude and code phase. Each bank of correlators, however, is similar to that of a brute-force acquisition, although possibly spread over a smaller range of uncertainty. This bank does not generate continuous numerically controlled oscillator (NCO) replicas of code and carrier phase. One of this chapter's contributions consists of a way to use these correlator banks as though they had produced continuous carrier and code NCO phases.

The accumulation measurements, produced by the bank of correlators, are

handled by optimal estimation techniques: batch nonlinear optimization and Kalman filtering. Batch estimation, in this case the Levenberg-Marquardt method, acts as a measurement pre-processor to the Kalman filter, as shown in Fig. 2.2. One could interpret its pre-processing calculations as a special coupled carrier-phase/carrier-Doppler-shift/code-phase discriminator.

Batch nonlinear optimization yields "pseudo-measurements" that the Kalman filter uses in the measurement update, while the main loop fulfills the ultimate purpose, which is to track the estimates of carrier phase, code phase and carrier amplitude that constitute the states of the dynamic model.

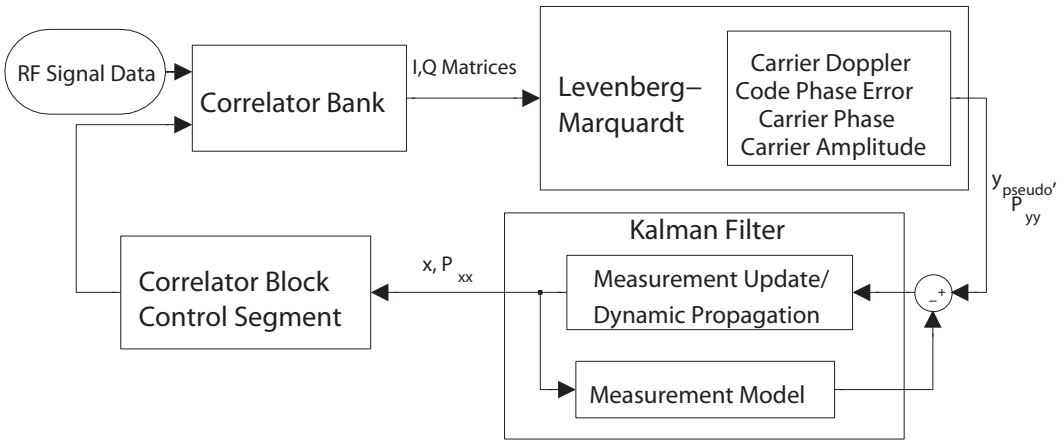

Figure 2.2: Block diagram of tracking algorithm, showing pre-processing behaviour of Levenberg-Marquardt batch estimation.

Batch fitting in Doppler-shift/code-phase space has also been investigated in Refs. [48] and [52]. Reference [48] uses batch correlators in both wide and narrow spans of Doppler-shift/code-phase uncertainty space, depending on the availability of previous measurements. It details a tracking mechanism where a search within a coarse grid is first performed to obtain rough estimates, but it does not specify any interpolation or estimation techniques for this step. Reference [52] also uses in-phase and quadrature accumulations (I's and Q's) from

multicorrelators in this two-dimensional space, but its method of deducing information uses two separate, ad-hoc quadratic fittings for its frequency and code delay discriminators, along with a traditional phase discriminator. It does, however, mention that this ad-hoc approach could be replaced by one based on an actual signal model, which is what is done in the present chapter. The present chapter also develops its phase discriminator in an optimal manner that is coupled to its fit procedure. Reference [52] uses the resulting discriminator outputs to drive a Kalman filter, while other Kalman filter tracking algorithms work directly with the I's and Q's as measurements [22, 51, 54]. The current chapter is similar to Ref. [52] in this respect.

Many other studies have already been conducted that test the use of multiple correlators, especially in the area of multipath mitigation techniques, but most only vary the code delay component of the replica signal with respect to the prompt signal. This is particularly useful for determining multipath effects due to their tendency to corrupt the correlation function. Multipath parameters can be found by curve-fitting with a method such as least-squares [9]. Reference [4] is another example of removing multipath by using many correlators in the code phase offset direction (more than 100) and by using a maximum-likelihood solution from a nonlinear curve fit around the peak of the correlation function. Reference [53] proposes an algorithm to quickly and efficiently compute a large number of correlator lags for similar purposes. NovAtel has built a Multipath Estimating Delay Lock Loop (MEDLL) receiver that uses many correlators with various lags and that estimates multipath components using the maximum likelihood criterion [46]. Reference [32] provides formulas for four types of software-based correlator output values, which depend on both Doppler frequency offset and code phase offset, but does not vary them in grid patterns that

span regions in Doppler-shift/code-phase space. The present chapter's proposed algorithm extends offsets in both Dopper and code phase directions and uses optimization methods designed to enable tracking with acquisition-like data, instead of to eliminate multipath.

The new algorithm has been evaluated by using measurements from a truth-model simulation of limb-scanning. These measurements simulate the accumulations from a given correlator bank. The simulated measurements account for the carrier Doppler shift and code phase offsets between the true limb-scanning signal and each replica associated with an element of the correlator bank.

This chapter begins by reviewing the signal model in Section II, from which the accumulation measurement model is derived in Section III. It then develops a dynamic model for carrier phase, code phase and carrier amplitude in Section IV. Sections V, VI and VII are devoted to the implementations of batch nonlinear optimization, the Kalman filter, and the PLL and DLL feedback control laws. In Sections VIII and IX, the chapter explains the mechanics of the limb scan simulation and analyzes the results of applying the algorithm to the simulation. Lastly, Section X is a brief discussion of the simulation results in a weak-signal case with random Markov-process dynamics.

2.3 Signal Model

This chapter's batch optimization and Kalman filter algorithms require a model of how carrier phase, carrier Doppler shift, carrier amplitude, and code phase are related to the correlations that are computed in a receiver. A correct accumulation model starts with a signal model that describes the output of the

RF front-end. This model assumes that the receiver RF front-end produces an intermediate-frequency output signal in the form

$$y(\tau_i) = Ad_kC[(\tau_i - t_{Sk})(1+\eta)]\cos[\omega_{IF}\tau_i + \phi_k + (\tau_i - t_{DLLk})\omega_{Davgk}] + n_i \qquad (2.1)$$

at sample time τ_i, where A is the signal amplitude, d_k is the 50 Hz GPS navigation data stream of ±1 values, $C(t)$ is the PRN code of the received signal, t_{Sk} is the PRN code start time of the received signal, $\eta = \omega_{Dk}/\omega_{carr}$ is the non-dimensional Doppler on the PRN code chipping rate, ω_{IF} is the nominal intermediate frequency, the frequency to which the RF front-end mixes the nominal carrier frequency ω_{carr}, ϕ_k is the k^{th} (negative) beat carrier phase, t_{DLLk} is start time of the kth accumulation interval, ω_{Davgk} is the average Doppler shift during the kth accumulation interval, and n_i is a sample of zero-mean discrete-time Gaussian white noise with variance σ_n^2. The carrier-to-noise density of the sampled signal is $C/N_0 = A^2/(4\sigma_n^2\delta\tau_S)$, assuming the noise bandwidth equals half the sampling frequency, where $\delta\tau_S = \tau_{i+1} - \tau_i$ is the RF front-end sampling interval. For the purposes of this dissertation, the navigation bit d_k is assumed to be known at all times. This assumption is relaxed in Section VI. In addition, note that ϕ_k is the time integral of the carrier Doppler shift, which makes it the negative of the usual definition of beat carrier phase. Also note that the DLL attempts to keep t_{DLLk} close to t_{Sk}. The DLL keeps track of t_{DLLk} as part of its strategy for providing a continuous record of its code phase estimates.

The navigation data bit d_k often can be known based on prediction, as in [31]. Data bit prediction relies on the repeatability of navigation messages. For example, the total GPS L1 navigation message is 12.5 minutes long and remains unchanged for periods of two hours. For a limb-scanning receiver in LEO, prediction could be based on bits recorded during a previous contact or on bits

uplinked from a ground receiver.

2.4 Accumulation Measurement Models

The receiver accumulates correlations between $y(\tau_i)$ and replicas of carrier and code signals. The recipes for its in-phase and quadrature accumulations take the form

$$I_k^{l,p} = \sum_{i=i_{k0}}^{i_{k0}+N_k-1} y(\tau_i) C[\tau_i - \Delta t_{DLLp,k} - t_{DLLk}] \cos[(\omega_{IF} + \omega_{PLLl,k})(i - i_{k0})\delta\tau_S + \phi_{l0k}] \quad (2.2)$$

$$Q_k^{l,p} = \sum_{i=i_{k0}}^{i_{k0}+N_k-1} y(\tau_i) C[\tau_i - \Delta t_{DLLp,k} - t_{DLLk}] \sin[(\omega_{IF} + \omega_{PLLl,k})(i - i_{k0})\delta\tau_S + \phi_{l0k}] \quad (2.3)$$

where the ranges of Doppler shifts and code delays that define the correlator bank are the following:

$$\omega_{PLLl,k} = [l - \frac{1}{2}(L-1)]\Delta\omega_{PLL} + \omega_{PLLavgk}; \qquad for\ l = 0, \ldots, L-1 \quad (2.4)$$

$$\Delta t_{DLLp,k} = [p - \frac{1}{2}(P-1)]\Delta t_{DLL} + \Delta t_{DLLavgk}; \qquad for\ p = 0, \ldots, P-1 \quad (2.5)$$

where $\Delta\omega_{PLL}$ and Δt_{DLL} are the carrier Doppler shift and code delay spacings of the correlator bank, and $\omega_{PLLavgk}$ and $\Delta t_{DLLavgk}$ are, respectively, the predicted carrier Doppler shift and code delay error for the interval. i_{k0} is the minimum i such that $t_{DLLk} \leq \tau_i$, N_k is the maximum N such that $\tau_{ik0+N-1} < t_{DLLk+1}$, and ϕ_{l0k} is the initial intermediate-frequency phase offset of the baseband mixing signal for the particular correlator and accumulation interval.

This model is different from traditional continuous-phase carrier NCO's, especially given that there are multiple NCO Doppler shifts. Equations (2.2) and (2.3) constitute recipes that will be implemented in an FPGA, or some other digital hardware, in order to calculate the accumulations for its bank of correlators.

The above model also provides a means of relating these accumulations to the signal parameters that the estimation methods will determine.

The following model is used to design estimators that deduce carrier phase, code phase, and carrier amplitude from the accumulation outputs of the bank of correlators. It has been constructed by substituting Eq. (2.1) into Eqs. (2.2) and (2.3), by using trigonometric product identities, by assuming that the summation will filter out frequencies near $2\omega_{IF}$, and by using approximations of nearly continuous-time sampling and large N_k. The final measurement model takes the form

$$\begin{aligned}\mathbf{h_k^{l,p}} &= \begin{pmatrix} I_k^{l,p} \\ Q_k^{l,p} \end{pmatrix} \\ &= \frac{AN_k}{2} d_k \begin{pmatrix} \cos(\phi_{PLLl,k} - \phi_{IFk} - \phi_{avgk}) \\ \sin(\phi_{PLLl,k} - \phi_{IFk} - \phi_{avgk}) \end{pmatrix} sinc[(\omega_{PLLl,k} - \omega_{Davgk})\delta t_k] R(\Delta t_{Smidk} - \Delta t_{DLLp,k}) \\ &\quad + \begin{pmatrix} n_{Ik}^{l,p} \\ n_{Qk}^{l,p} \end{pmatrix} \end{aligned} \tag{2.6}$$

where $\phi_{PLLl,k} = (\omega_{IF} + \omega_{PLLl,k})(\frac{N_k-1}{2})\delta\tau_s + \phi_{l0k}$ is the total Doppler-shifted mixing signals' intermediate-frequency phase at the midpoint of the accumulation samples, $\phi_{IFk} = mod\left\{\omega_{IF}[\tau_{ik0} + (\frac{N_k-1}{2})\delta\tau_s], 2\pi\right\}$ is the common-mode phase associated with the nominal intermediate frequency at the midpoint, ϕ_{avgk} is the average carrier phase over the accumulation interval, ω_{Davgk} is the average Doppler shift of the signal over the interval, and $\Delta t_{Smidk} = 0.5(\Delta t_{Sk} + \Delta t_{Sk+1})$ is the code phase error at the midpoint of the accumulation interval measured relative to the prompt PRN code. $\Delta t_{Sk} = t_{Sk} - t_{DLLk}$ is the code phase error at the start of the interval. If the accumulation interval is defined as $\delta t_{DLLk} = t_{DLLk+1} - t_{DLLk}$, then $\delta t_k = 0.5(\delta t_{DLLk} - \delta\tau_s)$ in the sinc functions of Eq. (2.6). $R(t)$ is the cross-correlation function of the PRN code replica with the received, distorted PRN

code. It is modelled with cubic splines at its slope discontinuities, in order to make its derivatives continuous and also take into account the actual rounding of the function's sharp corners due to the limited bandwidth of the RF front-end.

The vector $\mathbf{h_k^{l,p}}$ has two elements, but the correlator bank produces $P \times L$ such vectors. This entire set of correlation measurements can be stacked into the $2PL \times 1$ vector

$$\mathbf{h_k} = \begin{pmatrix} \mathbf{h_k^{1,1}} \\ \mathbf{h_k^{1,2}} \\ \vdots \\ \mathbf{h_k^{1,P}} \\ \mathbf{h_k^{2,1}} \\ \vdots \\ \mathbf{h_k^{L,P}} \end{pmatrix} \tag{2.7}$$

Similarly, the noise terms at the end of Eq. (2.6) can be stacked into the $2PL \times 1$ noise vector $\mathbf{n_k}$. The zero-mean, Gaussian discrete-time noise vector $\mathbf{n_k}$ is characterized by its $2PL \times 2PL$ noise covariance matrix, R_k. The necessary formulas for its elements are

$$\begin{aligned} E[(n_{Ik}^{l,p})(n_{Ik}^{j,q})] &= E[(n_{Qk}^{l,p})(n_{Qk}^{j,q})] \\ &= \frac{\sigma_n^2}{2} N_k \cos(\phi_{PLLl,k} - \phi_{PLLj,k}) sinc[(\omega_{PLLl,k} - \omega_{PLLj,k})\left(\frac{N_k - 1}{2}\right)\delta\tau_S] \\ &\quad \cdot R_{HB}(\Delta t_{DLLp,k} - \Delta t_{DLLq,k}) \end{aligned} \tag{2.8}$$

$$\begin{aligned} E[(n_{Ik}^{l,p})(n_{Qk}^{j,q})] &= -\frac{\sigma_n^2}{2} N_k \sin(\phi_{PLLl,k} - \phi_{PLLj,k}) sinc[(\omega_{PLLl,k} - \omega_{PLLj,k})\left(\frac{N_k - 1}{2}\right)\delta\tau_S] \\ &\quad \cdot R_{HB}(\Delta t_{DLLp,k} - \Delta t_{DLLq,k}) \end{aligned} \tag{2.9}$$

where $R_{HB}(t)$ is the auto-correlation of the PRN code replica with itself. The subscript $(.)_{HB}$ stands for high bandwidth, because $R(t)$ would equal $R_{HB}(t)$ if

the RF front-end had a high enough bandwidth. It is important to note that, although optimal estimation works best when the measurement model refers directly to the raw measurements and their errors, as in Eq. (2.1), in this case such an approach would be inefficient. The final measurement model in Eq. (2.6) and the related covariance matrix described in Eq.'s (2.8) and (2.9) retains most of the significant signal information if the correlator bank's carrier Doppler shifts and code delays are chosen wisely. The corresponding likelihood function becomes a good approximation of a likelihood function based directly on Eq. (2.1) because it effectively interpolates between accumulations on the correlator bank's grid.

2.5 Carrier Phase, Code Phase, and Carrier Amplitude Dynamics

The dynamics model for carrier phase assumes the form of three cascaded integrators driven by white nose:

$$\begin{pmatrix} \phi \\ \omega_D \\ \alpha \end{pmatrix}_{k+1} = \begin{pmatrix} 1 & \delta t_{DLLk} & \frac{1}{2}\delta t_{DLLk}^2 \\ 0 & 1 & \delta t_{DLLk} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \phi \\ \omega_D \\ \alpha \end{pmatrix}_k + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} w_{\phi k} \tag{2.10}$$

where $(\phi, \omega_D, \alpha)_k^T$ is the state vector containing carrier phase, carrier Doppler shift, and rate of change of carrier Doppler shift at the start of the k^{th} accumulation interval, or in other words, at time t_{DLLk}. Recall that δt_{DLLk} is the length of the k^{th} accumulation interval. $w_{\phi k}$, another zero-mean, discrete-time Gaussian white-noise sequence, is the carrier phase process noise.

The states of this linear system can be used to derive the average carrier

phase over the accumulation interval between times t_{DLLk} and t_{DLLk+1} [36]:

$$\phi_{avgk} = \begin{pmatrix} 1 & \frac{1}{2}\delta t_{DLLk} & \frac{1}{6}\delta t_{DLLk}^2 \end{pmatrix} \begin{pmatrix} \phi \\ \omega_D \\ \alpha \end{pmatrix}_k + \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} w_{\phi k} \tag{2.11}$$

This is the phase that is subtracted from the NCO phase in the measurement model expressed in Eq. (2.6). Similarly, the average Doppler shift over the accumulation interval is

$$\omega_{Davgk} = \begin{pmatrix} 0 & 1 & \frac{1}{2}\delta t_{DLLk} \end{pmatrix} \begin{pmatrix} \phi \\ \omega_D \\ \alpha \end{pmatrix}_k + \begin{pmatrix} \frac{1}{\delta t_{DLLk}} & 0 & 0 & 0 \end{pmatrix} w_{\phi k} \tag{2.12}$$

This is the Doppler shift that is subtracted from the NCO Doppler shift in the measurement model of Eq. (2.6).

The noise covariance matrix associated with the white process noise in this dynamic model takes into account the random walk acceleration of the line-of-sight (LOS) vector, as well as the random walks of the receiver clock frequency

and receiver clock phase. The covariance matrix for $w_{\phi k}$ is

$$E(w_{\phi k}w_{\phi k}^T) = q_{LOS}\begin{pmatrix} \frac{1}{20}\delta t_{DLLk}^5 & \frac{1}{8}\delta t_{DLLk}^4 & \frac{1}{6}\delta t_{DLLk}^3 & \frac{1}{72}\delta t_{DLLk}^5 \\ \frac{1}{8}\delta t_{DLLk}^4 & \frac{1}{3}\delta t_{DLLk}^3 & \frac{1}{2}\delta t_{DLLk}^2 & \frac{1}{30}\delta t_{DLLk}^4 \\ \frac{1}{6}\delta t_{DLLk}^3 & \frac{1}{2}\delta t_{DLLk}^2 & \delta t_{DLLk} & \frac{1}{24}\delta t_{DLLk}^3 \\ \frac{1}{72}\delta t_{DLLk}^5 & \frac{1}{30}\delta t_{DLLk}^4 & \frac{1}{24}\delta t_{DLLk}^3 & \frac{1}{252}\delta t_{DLLk}^5 \end{pmatrix}$$
$$+ S_g\omega_{carr}^2\begin{pmatrix} \frac{1}{3}\delta t_{DLLk}^3 & \frac{1}{2}\delta t_{DLLk}^2 & 0 & \frac{1}{8}\delta t_{DLLk}^3 \\ \frac{1}{2}\delta t_{DLLk}^2 & \delta t_{DLLk} & 0 & \frac{1}{6}\delta t_{DLLk}^2 \\ 0 & 0 & 0 & 0 \\ \frac{1}{8}\delta t_{DLLk}^3 & \frac{1}{6}\delta t_{DLLk}^2 & 0 & \frac{1}{20}\delta t_{DLLk}^3 \end{pmatrix}$$
$$+ S_f\omega_{carr}^2\begin{pmatrix} \delta t_{DLLk} & 0 & 0 & \frac{1}{2}\delta t_{DLLk} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{1}{2}\delta t_{DLLk} & 0 & 0 & \frac{1}{3}\delta t_{DLLk} \end{pmatrix} \tag{2.13}$$

where q_{LOS} is the power spectral density of the continuous-time white noise that drives the acceleration random walk, S_g is the power spectral density of the white noise that drives the receiver clock frequency random walk, and S_f is the power spectral density of the white noise that drives the clock phase random walk [36].

The dynamic model for the PRN code phase keeps track of the true code start and stop times associated with the nominal PRN code segment for the accumulation interval:

$$t_{Sk+1} = t_{Sk} + \frac{\omega_{carr}\delta t_{nom} - \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} w_{\phi k}}{\omega_{carr} + \omega_{Dk} + \frac{1}{2}\delta t_{nom}\alpha_k} + w_{tSk} \tag{2.14}$$

where δt_{nom} is the nominal length of the code segment and w_{tSk} is a white noise term that models the random walk of code phase. The second term on the right-hand side of Eq. (2.14) is the carrier-aiding term that captures the coupling between carrier Doppler shift and code chipping rate. This carrier aiding term

matches the true number of carrier cycles in the accumulation interval with the nominal number that are broadcast. The dynamic model for code phase error is therefore

$$\Delta t_{Sk+1} = \Delta t_{Sk} + \frac{\omega_{carr}\delta t_{nom} - \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} w_{\phi k}}{\omega_{carr} + \omega_{Dk} + \frac{1}{2}\delta t_{nom}\alpha_k} - \delta t_{DLLk} + w_{tSk} \tag{2.15}$$

Note that the second and third terms on the right-hand side of Eq. (2.15) comprise the difference between the true length of the code segment and the DLL's estimate of it. The DLL attempts to keep this difference near zero. The PRN code phase error at the midpoint of the accumulation, which is needed as part of the argument for the PRN code autocorrelation function in Eq. (2.6) is

$$\Delta t_{Smidk} = \Delta t_{Sk} + \frac{1}{2}\frac{\omega_{carr}\delta t_{nom} - \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} w_{\phi k}}{\omega_{carr} + \omega_{Dk} + \frac{1}{2}\delta t_{nom}\alpha_k} - \frac{1}{2}\delta t_{DLLk} + \frac{1}{2}w_{tsk} \tag{2.16}$$

Also modeled as a purely random walk is carrier amplitude:

$$A_{k+1} = A_k + w_{Ak} \tag{2.17}$$

where w_{Ak} is the discrete-time white process noise that drives the random walk. This equation can be used to deduce the average amplitude of the accumulations:

$$A_{IQk} = \frac{1}{2}N_k A_k + \frac{1}{4}N_k w_{Ak} \tag{2.18}$$

2.6 Batch Nonlinear Optimization Using the Levenberg-Marquardt Method

Kalman filters typically contain two stages of computation: Dynamic propagation and measurement update. Measurement updating adjusts the a priori state

estimate based on incoming measurements. Batch optimization in the present algorithm provides these measurements as a multi-correlator vector discriminator of carrier Doppler shift, code phase, carrier phase, and carrier amplitude.

Specifically, batch optimization fits accumulation data coming out of the correlator bank to the measurement model in Eq. (2.6) and yields the most likely signal parameters associated with the best fit to the model. These parameters are the batch filter unknowns for which the Levenberg-Marquardt algorithm seeks estimates. The unknowns are average Doppler shift, midpoint code phase error, average carrier phase, and average carrier amplitude, denoted in this chapter by $(\omega_{Davgk}, \Delta t_{Smidk}, \phi_{avgk}, A_{IQk})$. The measurement models that relate these variables to the Kalman filter states $(\phi_k, \omega_{Dk}, \alpha_k, \Delta t_{Sk}, A_k)$ were given in the previous section. In essence, the output of the batch optimization algorithm provides something akin to partial measurement linearization and sensitivity adjustment of the accumulations about the optimal values of these signal parameters.

Batch nonlinear optimization starts by choosing the correlator, indexed by l, p, that has the highest $I^2 + Q^2$ accumulation power, along with its nearest neighbours. This search is conducted along both the Doppler shift and code phase directions, and the calculations are the same as those in a normal GNSS acquisition search. Figure 2.3 shows a superposition of continuous, theoretical power, which the receiver never actually sees, and discrete, correlator-measured power.

Each dot in Fig. 2.3 represents a correlator that mixes the RF front-end output signal with the appropriate NCO Doppler shift and code phase offset within the specified ranges. Note that the effects of noise have been neglected in generating Fig. 2.3, thereby causing the dots to fall exactly on the theoretical plot. The

number of correlators used will depend on the uncertainty of these two signal parameters in the current accumulation interval. Uncertainty here is based on the Doppler shift and code phase offset variances in the state error covariance matrix determined by the Kalman filter. The spacing of the correlator grid is predetermined in the simulation with reasonable values that are close enough to find the peak, but not so close that they begin to cause excessive computational cost and numerical conditioning problems.

The red dots depict the correlators that are chosen to have their accumulation measurements sent to the batch estimation algorithm for data fitting. A clearer view of the correlator grid and these selected points is shown in the power contour plot of Fig. 2.4.

In the second plot, it is obvious that none of the chosen correlators, indicated in red, need to be exactly on the peak of the power function. The Levenberg-Marquardt solution, however, will give the best accuracy if the selected points span the peak. In this simulation, the nearest neighbours are chosen such that they are within 1 code chip or one code delay spacing Δt_{DLL}, and one accumulation frequency or one Doppler shift spacing $\Delta\omega_{PLL}$, whichever ones are larger, of the values that gave the peak accumulation power.

Data fitting is performed by minimizing the cost function

$$J(\omega_{Davgk}, \Delta t_{Smidk}, \phi_{avgk}, A_{IQk}) = \frac{1}{2}[\mathbf{y_{trk}} - \mathbf{h_{trk}}(\omega_{Davgk}, \Delta t_{Smidk}, \phi_{avgk}, A_{IQk})]^T \cdot R_{trk}^{-1}[\mathbf{y_{trk}} - \mathbf{h_{trk}}(\omega_{Davgk}, \Delta t_{Smidk}, \phi_{avgk}, A_{IQk})] \quad (2.19)$$

The measurement vector $\mathbf{y_{trk}}$ is a truncated version of the vector of I's and Q's obtained directly from the correlator bank, and the vector $\mathbf{h_{trk}}$ is the corresponding truncated version of h_k from Eq. (2.7). This truncation eliminates the corre-

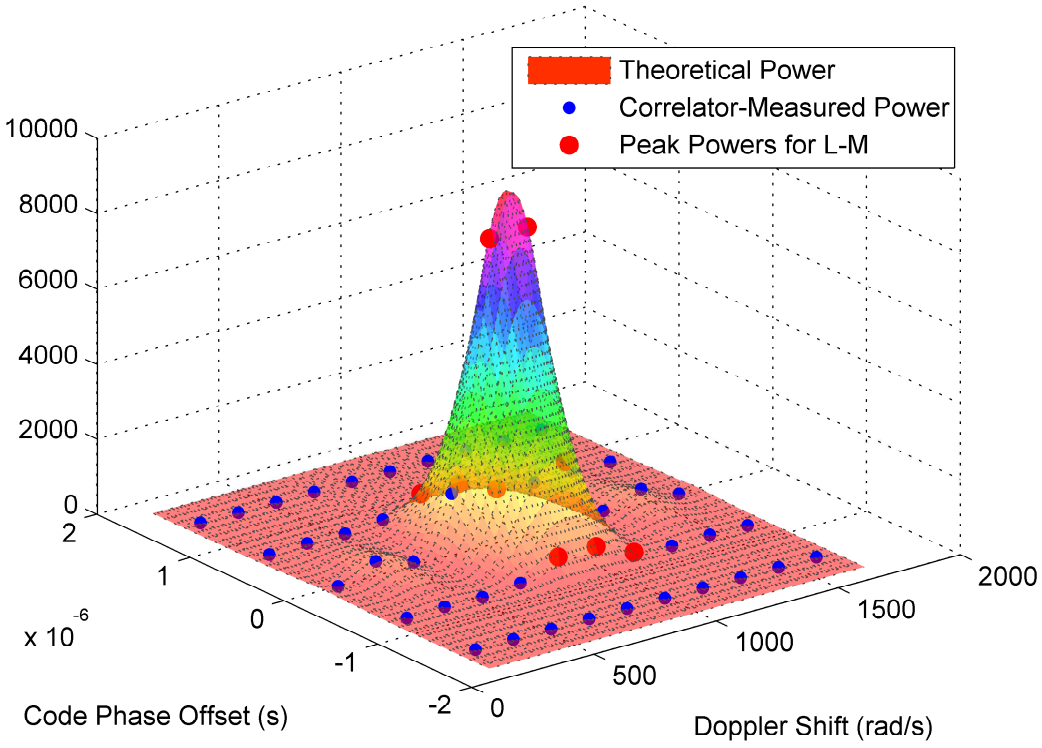

Figure 2.3: Measured power of accumulations from a correlator bank, superimposed on theoretical power.

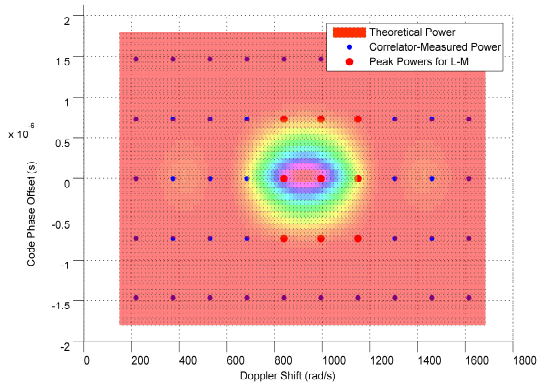

Figure 2.4: Power contours mapped onto a bank of correlators.

lations associated with the blue dots in Figs. 2.3 and 2.4, to retain only those associated with the red dots, i.e. the correlations that have significant power. The inverse covariance matrix R_{trk}^{-1} will likewise be a smaller version of the original

noise covariance matrix. The cost J is the negative log likelihood of its four input parameters. Therefore, the Levenberg-Marquardt solution is the maximum-likelihood solution.

Before implementing the Levenberg-Marquardt algorithm, it is helpful to recognize that ϕ_{avgk} and A_{IQk} can be optimized analytically. In order to do this, define

$$\mathbf{x} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} = \begin{bmatrix} (A_{IQk}\cos(\phi_{avgk}) & A_{IQk}\sin(\phi_{avgk}) \end{bmatrix} \tag{2.20}$$

and set the derivative of the above cost function, with respect to this vector $\mathbf{x}$, to zero. The resulting necessary condition is linear in $\mathbf{x}$, and can be solved exactly. Analytically optimized values of A_{IQk} and ϕ_{avgk} can be calculated for every pair of ω_{Davgk} and Δt_{Smidk}:

$$A^*_{IQk} = \|\mathbf{x}\| \tag{2.21}$$

$$\phi^*_{avgk} = atan2(x_2, x_1) \tag{2.22}$$

Due to its structure, the measurement model can be easily rearranged so that this can be done. With analytic optimization, the Levenberg-Marquardt algorithm will only need to search in the two directions of Doppler shift and code phase error, instead of in four directions. This cuts down much of the computational expense.

The Levenberg-Marquardt implementation used here has two special features [17]. The first is its use of the exact cost function Hessian, rather than the approximate Hessian that uses only first derivatives of the measurement function $\mathbf{h_{trk}}$. The second feature is a modification of how the algorithm calculates the parameter that limits the step size. In addition to the usual requirement of cost decrease, the parameter can be adjusted upward to compensate for an indefinite cost Hessian and to keep the solution within the bounds of the chosen

points in Doppler-shift/code-phase space. This limitation, paired with the use of the peak power point as the first guess, tends to guarantee that the solution is the global minimum.

The final optimized batch states become the Kalman filter's measurements. These have been called "pseudo-measurements" in the introduction due to the fact that they do not correspond directly to the measurement model in Eq. (2.6). Instead, the Kalman filter compares these "measurements" of (ω_{Davgk}, Δt_{Smidk}, ϕ_{avgk}, and A_{IQk}) to the models established in Eqs. (2.12), (2.16), (2.11) and (2.18), respectively. There is also an error covariance matrix associated with this measurement model. It is set equal to the inverse of the Levenberg-Marquardt optimal cost Hessian. Although this is more like a Cramer-Rao lower bound, it is a reasonable choice. The Cramer-Rao lower bound is the right choice in the limit of many samples [2]. Reference [49] gives a maximum likelihood estimate of (differential) Doppler and delay for a generalized ambiguity function, along with a derivation of its Cramer-Rao bound.

Note that an alternate method would be to apply a nonlinear Kalman filter directly to the accumulation measurements described in Eq. (2.6), but the present method has the advantage of increased accuracy by "linearizing" about the parameters that give peak accumulation power, instead of only linearizing about the Kalman filter's a priori estimate. The Levenberg-Marquardt algorithm is also a way to transition between the large collections of accumulations produced by the correlator bank to a reasonable set of measurements that the Kalman filter can easily incorporate.

2.7 Implementation of Kalman Filter

The Kalman filter used for the signal tracking algorithm is a nonlinear extended square root information filter (EKF). The Kalman filter makes use of the dynamic models for carrier phase, carrier Doppler shift, rate of change of carrier Doppler shift, code phase error, and carrier amplitude. This model is a combination of Eqs. (2.10), (2.15) and (2.17), and it is used in dynamic propagation. The model used for measurement update is the output of the batch optimization, as described in the previous section. Both processes of dynamic propagation and measurement update are nonlinear due to carrier-aiding being present in code phase dynamics. This weak nonlinearity is the only one remaining after the Levenberg-Marquardt solution produces the "measurements".

Also implemented within the Kalman filter are four data tests that are used in order to decide whether or not to perform a measurement update. All four tests must be passed in order to perform an update. The first test accepts data only if the peak accumulation power in the correlator bank does not lie on a boundary in the Doppler shift or code delay range. The second test requires that the Levenberg-Marquardt algorithm converges in a reasonable number of iterations. The third test examines the optimal Leveberg-Marquardt cost. It should be half of a chi-squared sample from a distribution of degree $N_{trk} - 4$, where N_{trk} is the number of elements in $\mathbf{y_{trk}}$ and in $\mathbf{h_{trk}}$, i.e., twice the number of selected red dots in Figs. 2.3 and 2.4. If the optimal Levenberg-Marquardt cost is too high, say high enough that the random probability of generating it is less than 10^{-4}, then the measurement is rejected. The fourth test performs the Kalman filter measurement update, and examines the value of the sum of the squares of the normalized innovation vector. It should be a sample from a

chi-squared distribution of degree four. If the value is too high, then the measurement update is rejected. Such rejections, however, are rare, except at very low carrier-to-noise ratios.

The Kalman filter also handles full-cycle phase ambiguities by adding multiples of full cycles to the carrier phase measurement provided by the batch filter:

$$\tilde{\phi}_{avgk} = \phi_{avgk} + 2\pi round\left(\frac{\bar{\phi}_{avgk} - \phi_{avgk}}{2\pi}\right) \tag{2.23}$$

where $\tilde{\phi}_{avgk}$ is the actual phase in radians used in the measurement update, ϕ_{avgk} is the Levenberg-Marquardt solution, and $\bar{\phi}_{avgk}$ is the Kalman filter's a priori expected phase calculated from the right-hand side of Eq. (2.11). This forces the filter's phase innovation to be less than π. The rounding procedure can be modified to force the innovation to be less than $\pi/2$ instead, in order to deal with half-cycle ambiguities that would arise due to unknown navigation data bits. This change would be needed if, contrary to the assumption of Section II, the data bits could not be predicted.

2.8 PLL and DLL Feedback Control Laws

Feedback control laws are needed in order to complete the main loop shown in Fig. 2.2, by implementing the correlator block control segment. These laws use the most recent Kalman filter a posteriori state estimates in order to predetermine the accumulation intervals, the NCO carrier Doppler shifts, as defined in Eq. (2.4), and the NCO code phase offsets, as defined in Eq. (2.5). The feedback from a given accumulation interval is used to set these values two accumulation intervals forward.

The DLL feedback control law determines the $k + 2^{nd}$ accumulation interval

as follows:

$$\delta t_{DLLk+2} = \Delta\hat{t}_{Sk+1} + \frac{\omega_{carr}\delta t_{nom}}{\omega_{carr} + \hat{\omega}_{Dk+1} + 0.5\delta t_{nom}\hat{\alpha}_{k+1}} + \frac{\omega_{carr}\delta t_{nom}}{\omega_{carr} + \hat{\omega}_{Dk+1} + 1.5\delta t_{nom}\hat{\alpha}_{k+1}} - \delta t_{DLLk+1} \tag{2.24}$$

where $\hat{\omega}_{Dk+1}$, $\hat{\alpha}_{k+1}$, and $\Delta\hat{t}_{Sk+1}$ are a posteriori state estimates at time t_{DLLk+1}. The effect of this control law is to use the code phase offset estimate $\Delta\hat{t}_{Sk+1}$ and the predicted true lengths of the $k+1^{st}$ and $k+2^{nd}$ accumulation intervals to choose δt_{DLLk+2} in a way that causes the predicted code phase offset at t_{DLLk+3} to be zero. The principle behind this choice is that of trying to keep the centre of the batch of code delays and the centre of the accumulation interval aligned with the best estimate of the prompt code interval.

The DLL also needs to predict the expected average code phase error for the $k+2^{nd}$ accumulation interval. Although the expected error is zero at the end of the interval, as per the design of Eq. (2.24), it is not necessarily zero at the beginning. In fact, it equals half the sum of the first, second and fourth terms on the right-hand side of Eq. (2.24):

$$\Delta t_{avgk+2} = \frac{1}{2}(\Delta\hat{t}_{sk+1} + \frac{\omega_{carr}\delta t_{nom}}{\omega_{carr} + \hat{\omega}_{Dk+1} + 0.5\delta t_{nom}\hat{\alpha}_{k+1}} - \delta t_{DLLk+1}) \tag{2.25}$$

This predicted value is used as $\Delta t_{DLLavgk+2}$ in the $k+2$ version of Eq. (2.5). Equation (2.5) also requires a choice of the PRN code spacing Δt_{DLL} and the number of code offsets P. Δt_{DLL} is normally chosen in the range 0.5 to 1 code chip lengths. P is normally chosen so that $P\Delta t_{DLL}$ is four to six times the Kalman filter's 1-σ code phase uncertainty, as computed from its covariance matrix. Note that a minimum value of $P = 3$ is enforced when the code delay uncertainty is very small.

The nominal Doppler shift predicted by the PLL for the $k+2^{nd}$ accumulation

interval is

$$\omega_{avgk+2} = \hat{\omega}_{Dk+1} + (\delta t_{DLLk+1} + \frac{1}{2}\delta t_{DLLk+2})\hat{\alpha}_{k+1} \tag{2.26}$$

This predicted value is used as $\omega_{PLLavgk+2}$ in the $k + 2$ version of Eq. (2.4). This PLL law is derived to keep the centre of the batch of Doppler shift values aligned with the best estimate of Doppler shift at the centre of the accumulation interval. Equation (2.4) also requires a choice of the Doppler shift spacing $\Delta\omega_{PLL}$ and the number of Doppler offsets L. $\Delta\omega_{PLL}$ is normally chosen in the range $\pi/\delta t_{nom}$ to $2\pi/\delta t_{nom}$, i.e. from $1/2$ to 1 times the accumulation frequency. L is normally chosen so that $L\Delta\omega_{PLL}$ is four to six times the Kalman filter's 1-σ Doppler shift uncertainty. Note, however, that a minimum value of $L = 3$ is always used, even when the Doppler shift uncertainty is very small.

The "hat" accents denote estimates given by the Kalman filter, for time t_{DLLk+1}. These do not become available until shortly after that time, because they are based on accumulations from the interval that ends at that time. Presumably, the receiver's processor can finish computing all of these feedback quantities during the interval from t_{DLLk+1} to t_{DLLk+2}, so that they will be available to the correlator block control segment during the interval from t_{DLLk+2} to $t_{DLLk+3} = t_{DLLk+2} + \delta t_{DLLk+2}$.

2.9 Limb Scan Simulation

The rising-GPS-satellite limb-scanning application, with a LEO satellite carrying a GPS receiver, is an ideal use for this acquisitionless tracking system. The most useful meteorological data occur just at the initial time when the satellite rises,

but it is difficult to start tracking the signal immediately with traditional receiver algorithms. An acquisition would waste some of this valuable data.

To investigate the new system's performance on this problem, a high-fidelity truth-model simulation of a limb scan has been developed. Its goal is to show that the new Kalman Filter can start tracking the signal with high accuracy on its carrier and code phase immediately upon availability. Thus, meteorological data could be captured for limb scans with minimum altitudes of only several meters.

The truth-model simulation includes LEO orbital dynamics, namely, a circular orbit with 700 km of altitude and 98.2° of inclination. The GPS satellite orbital dynamics are modelled using standard GPS ephemerides and orbit calculations. As the orbital dynamics cause the signal to penetrate the atmosphere, the signal path experiences refraction dictated by the generalized 3D Snell's law differential equation. The differential equation is applied through modelled neutral atmosphere density and ionosphere electron density distributions. The solution for the refracted path is that of a two-point boundary value problem (TPBVP) between each LEO and GPS satellite location. Figure 2.5 renders an example of a curve determined by the TPBVP solver. The plot shows the geometry of the Earth limb (blue), the straight-line path (green), which is occulted, and the refracted bent path (red), which has a minimum altitude of 54 m.

The effects on carrier Doppler shift and pseudorange are calculated at discrete time points and then interpolated between these times. These effects are due to tropospheric and ionospheric refraction, and they cause large initial perturbations in pseudorange and Doppler shift, as shown in the example of Fig. 2.6. The modelled refraction includes the phenomena of geometric

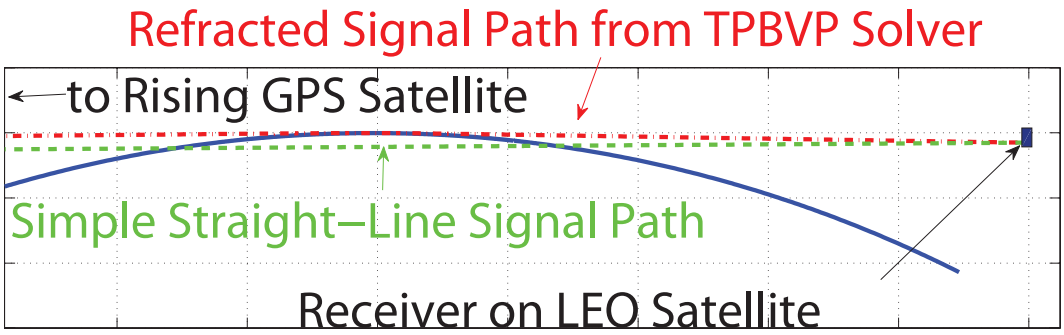

Figure 2.5: Sketch of limb-scan geometry.

path bending, signal group-delay/phase-delay in the troposphere, and group-delay/phase-advance in the ionosphere. The change in Doppler shift in the first

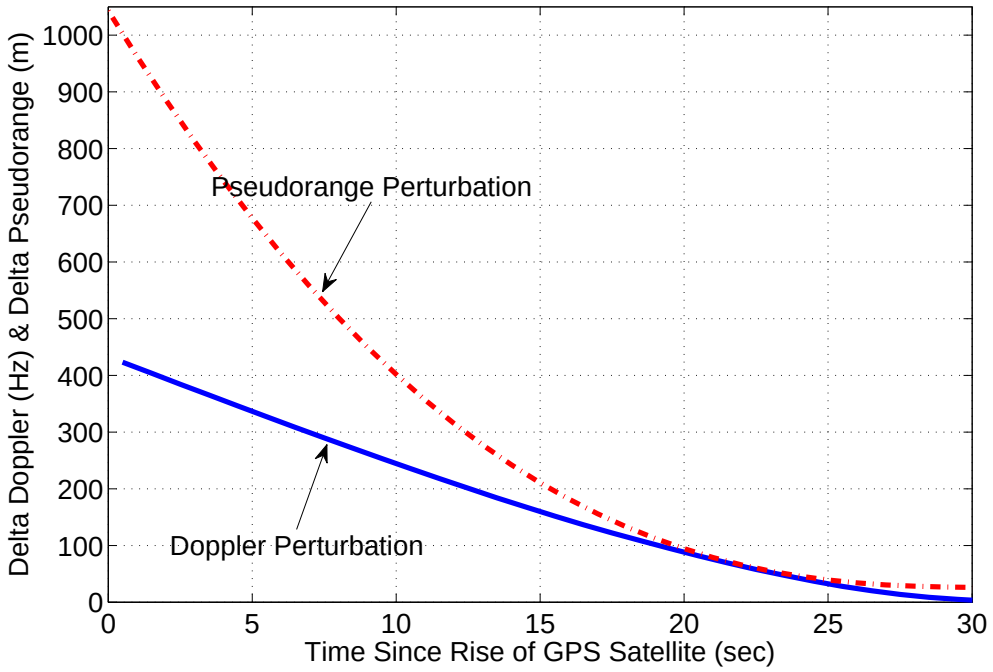

Figure 2.6: Pseudorange and carrier Doppler shift perturbation time histories due to refraction.

20 seconds is about 300 Hz, translating into an apparent unmodelled acceleration of about 0.3 g at the L1 carrier frequency. The initial pseudorange offset is 1040 m, which implies a code phase error of about 3.5 chips of C/A code.

The effects of receiver clock error are also simulated using a standard Markov model. It is given in Ref. [7].

2.10 Simulation Test Results

The following are the results of applying the new algorithm to the limb-scan truth-model simulation. The overall simulation has been implemented in MATLAB and uses correlator grid spaces of $\Delta\omega_{PLL}$ = 37.5 Hz in Doppler shift and Δt_{DLL} = 0.75 chips in code phase offset. These are reasonable, albeit ad-hoc, values that are large enough to avoid numerical issues such as poor conditioning of noise covariance matrices. The assumed sampling interval is 175 ns and the RF bandwidth is 3 MHz. The receiver clock model uses the parameters $h_0 = 10^{-22}s$ and $h_{-2} = 7.6 \times 10^{-24}s^{-1}$, as defined in Ref. [7]. These clock parameters yield a minimum root Allen variance of 10^{-11} at a delay of 1 s, consistent with an ovenized crystal oscillator, which is sometimes used in spacecraft applications. Note, however, that parameters consistent with a temperature-compensated crystal oscillator have been tested to yield similar results. q_{LOS} is set to 120 rad^2/s^5; this value is based on measures of the degree of error between "truth" phase and the rate at which it deviates from a quadratic in the limb-scan time histories. The nominal model accumulation interval δt_{nom} is set to 0.02 s, giving 50 Hz accumulations. The carrier frequency ω_{carr} corresponds to that of the L1 signal.

Figure 2.7 plots the simulated "truth" time history of carrier Doppler shift along with the Kalman filter estimate of this quantity, from the point when the GPS satellite rises above the Earth limb as viewed by the LEO satellite. The

Kalman filter obviously converges rapidly to a very good Doppler shift estimate. As shown in the lower plot of the figure, the root mean square error for Doppler shift in the part of the 120 s interval after convergence is 0.0423 Hz. The Kalman filter's average σ value for this same interval of steady-state tracking is 0.0699 Hz. It is important to consider not only the speed with which the PLL achieves lock, but also its successful pull-in from an initial Doppler shift offset of 750 Hz, despite using 50 Hz accumulations. These conditions would prevent a traditional PLL with a 1-, 2-, or 3-correlator discriminator from ever achieving lock.

Similar results are seen for PRN code phase error in Fig. 2.8. The DLL also achieves and maintains lock very quickly, even with an initial code phase error of 7 chips. The root mean square error during the interval of steady-state tracking is 0.0064 C/A code chips. Uncertainty is plotted in the form of the Kalman filter's σ values, which drop to 0.0081 chips, on average, after only about 0.02 s.

The simulated "truth" carrier-to-noise ratio and its estimate are shown in Fig. 2.9. The signal starts with about 15 dB of attenuation due to atmospheric path loss and the effects of refractive lensing. This power loss disappears after 60 s, by which time the signal path is largely clear of the atmosphere. The congruence of the "truth" and estimated C/N_0 plots implies that signal amplitude is also being tracked effectively.

The final signal parameter of interest is carrier phase. Its error time history is graphed in Fig. 2.10. Note once again that within about three accumulation intervals, the phase error is pulled in from half a cycle to a few degrees, giving a root mean square error of 0.0037 cycles during the steady-state portion of the

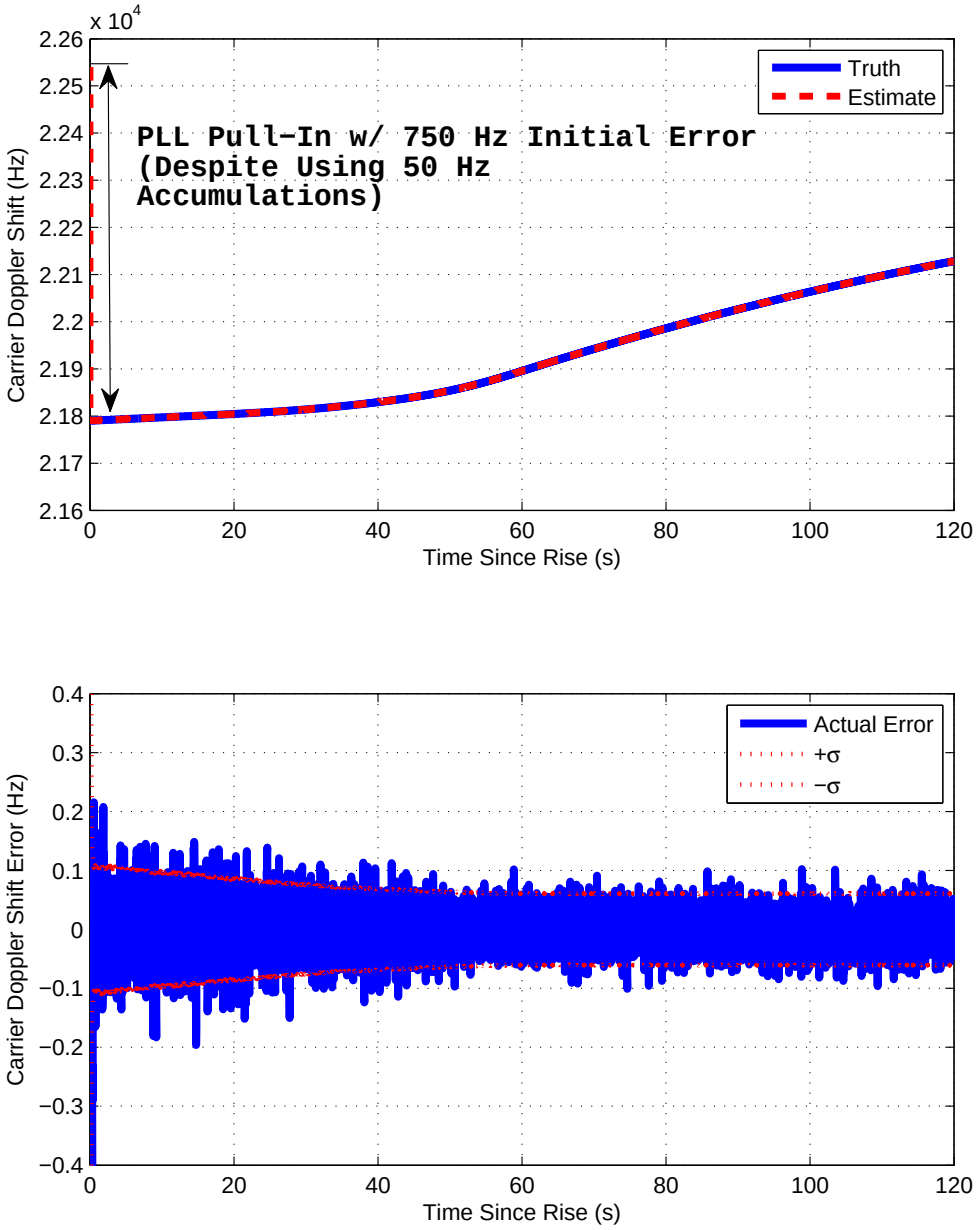

Figure 2.7: Top: Time histories of "truth" and estimated carrier Doppler shift. Bottom: Magnified time history of Doppler shift error and its 1-σ bounds.

simulation. The Kalman filter's $\pm\sigma$ uncertainty is effectively infinite at filter initialization, but it decreases to a steady-state average of 0.0037 cycles after the

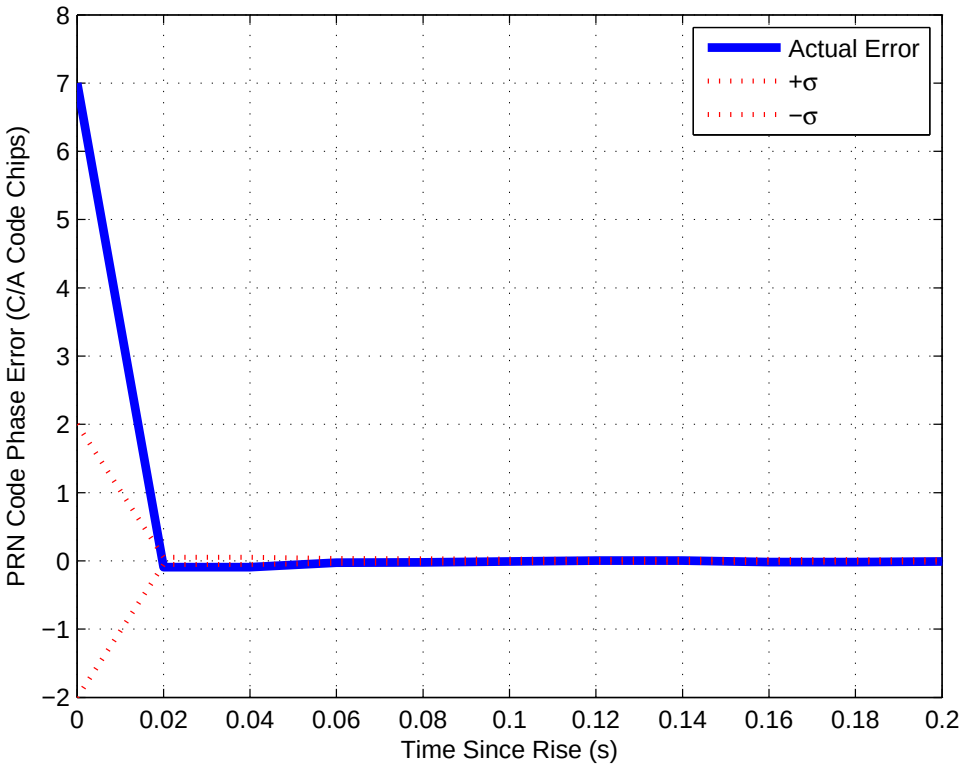

Figure 2.8: Time histories of PRN code phase error and its 1-σ bounds.

first few accumulation intervals. The algorithm responds to the aforementioned drops in uncertainty by lowering the number of correlators that must be used, saving computational cost. Time histories of these numbers are shown in Fig. 2.11. The bank of correlators starts with $L = 45$ NCO Doppler shifts and $P = 17$ different NCO code phase offsets.

The rapid reductions in uncertainty, as indicated by the Kalman filter covariance matrix, translate into drops in the numbers of NCO grid points both in the Doppler shift offset direction and in the code offset direction. After a short time, only three correlators are used in each direction in order to maintain lock.

The simulation was also successfully tested on the L2 signal frequency, as

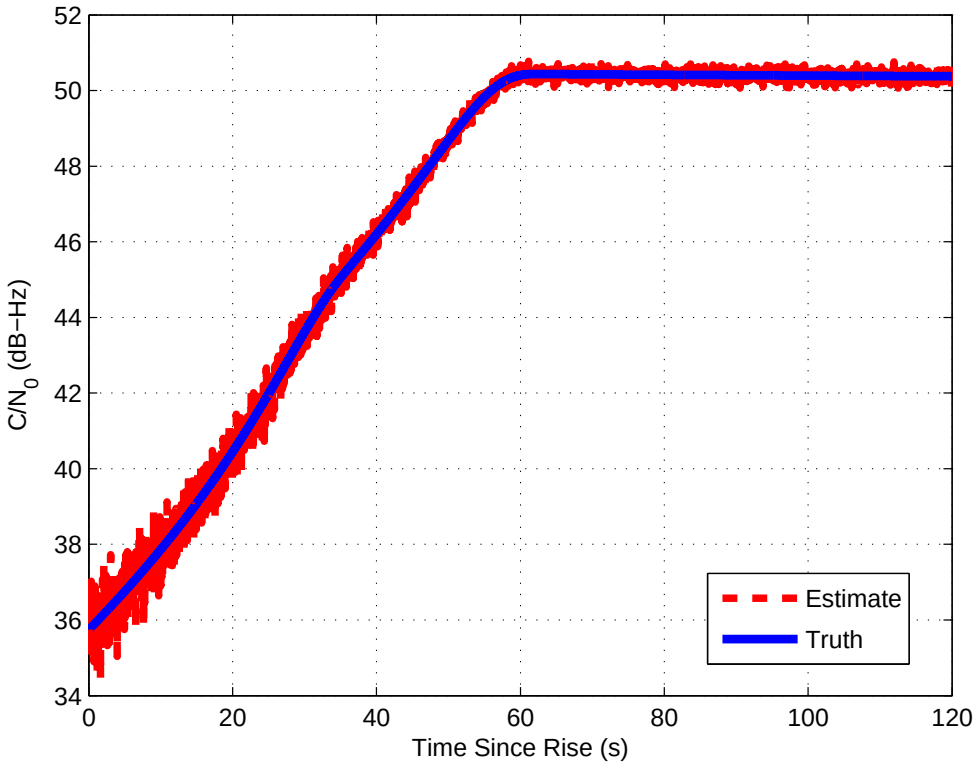

Figure 2.9: Time histories of "truth" and estimated carrier-to-noise ratios.

well as with P(Y) code, which has larger uncertainties than the C/A code in terms of chip ranges associated with pseudorange perturbations. Theoretically, this PLL/DLL combination has infinite code-offset and Doppler-shift pull-in ranges, if supplied with an infinite number of correlators and if computational overhead is not an issue. The algorithm will always achieve lock, as long as the uncertainties are initialized such the the first correlator bank of Fig. 2.1 captures the true values so that a valid correlation peak is seen by the receiver.

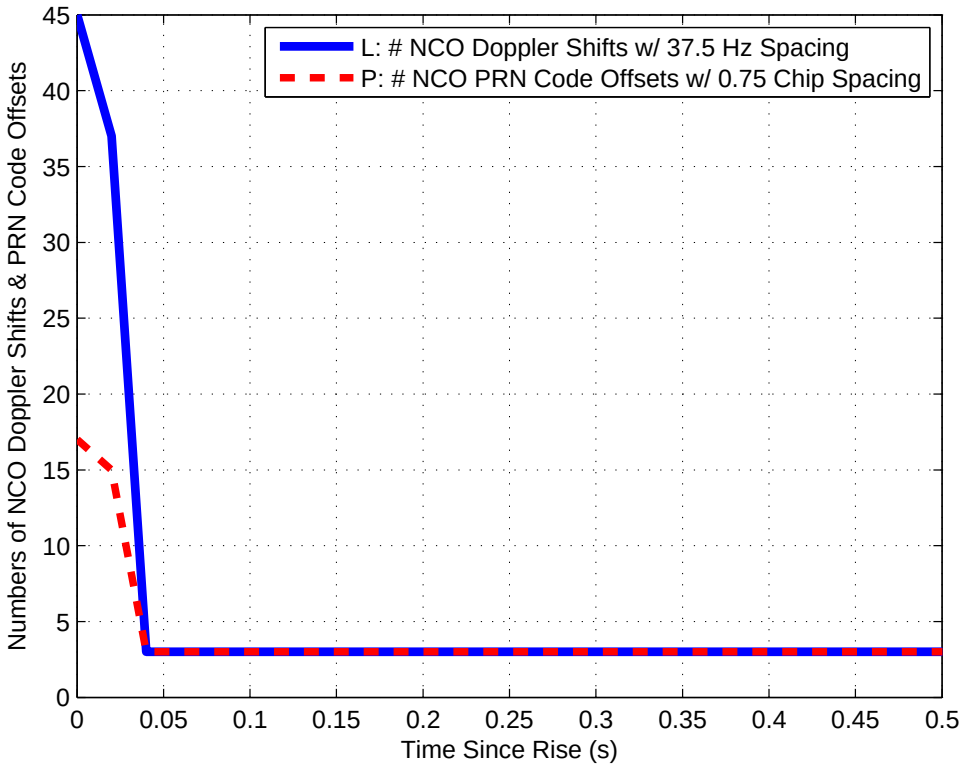

Figure 2.10: Time histories of carrier phase error and its 1-σ bounds.

2.11 Weak Signal Tracking

The tracking algorithm has also been tested on weak signals. Instead of using a limb scan simulation, the "truth" dynamics model used for this case is a simulation of the Markov models in Eqs. (2.10), (2.15), and (2.17), and the "truth" measurements are simulations of Eqs. (2.11), (2.12), (2.16), (2.18), and (2.6). In this application, the system can achieve and maintain lock down to about $C/N_0 = 25$ dB-Hz. This threshold might also be a few dB-Hz higher if navigation data bits were unknown because the ambiguity resolution in Eq. (2.23) would need to be modified to remove half-cycle ambiguities instead of full-cycle ambiguities. However, estimation of the amplitude state is consistently higher than the true

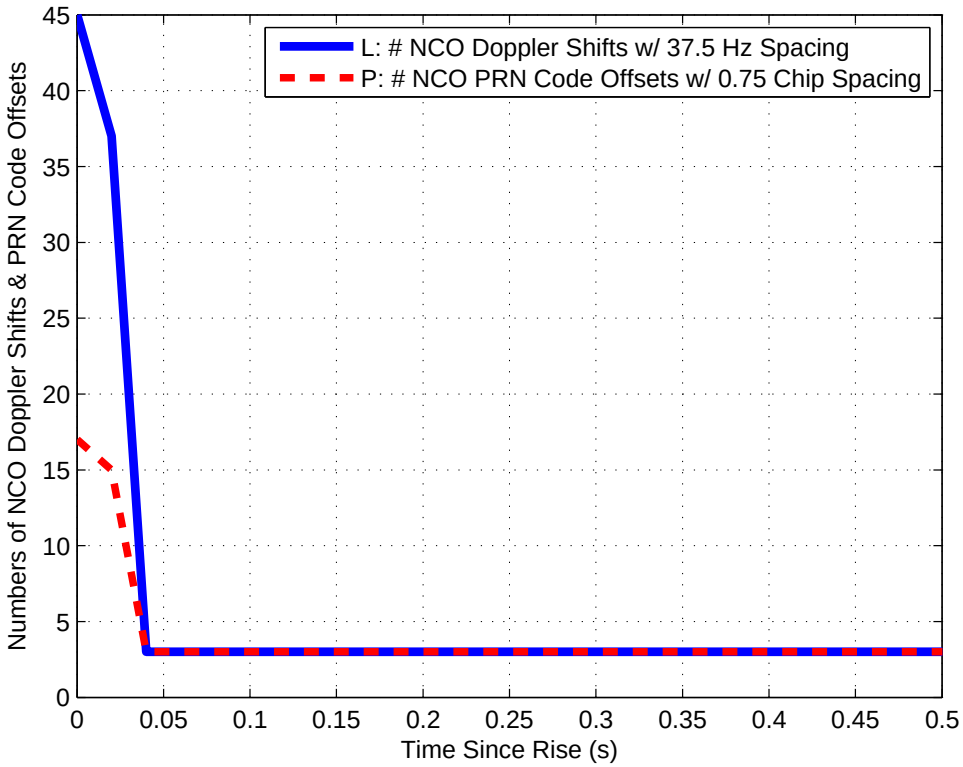

Figure 2.11: Time history of correlator bank size parameters.

values. This offset increases with lower carrier-to-noise ratios, and it arises because the maximum-likelihood estimator of the Levenberg-Marquardt solution is biased, and because of the systematic choice of the correlations with the highest power. Figure 2.12 shows this behaviour of the estimated carrier-to-noise ratio, which occurs even if lock on all other states is maintained, as shown in Figs. 2.13 and 2.14, for a signal of $C/N_0 \cong 27$ dB-Hz. Note that the Doppler shift magnitudes used for this simulation correspond to those typically seen by a receiver onboard a LEO satellite. Further work must be done, perhaps to more closely approximate a Bayesian nonlinear filter, in order to track very weak signals, i.e., $C/N_0 < 25$ dB-Hz, with this type of PLL/DLL. The required major changes might be made along the lines of Ref. [36], but they are beyond the

scope of this dissertation. Of course, no method will work for very weak signals if the bandwidth needs to be high and if there is no aiding available, such as from an IMU.

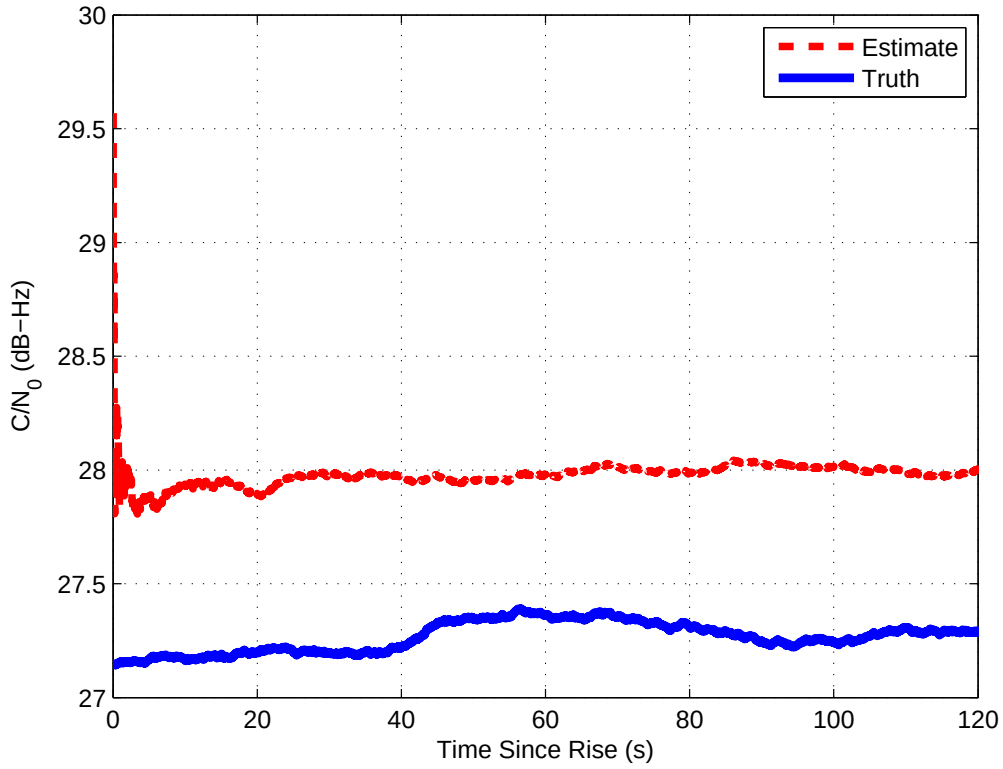

Figure 2.12: Time histories of "truth" and estimated carrier-to-noise ratios for a weak signal.

2.12 Conclusion

A new GPS signal tracking algorithm has been developed to combine PLL and DLL functions with accumulation data from an acquisition-like correlator bank. Its purpose is to achieve rapid tracking with large pull-in regions without the need for a separate initial acquisition. The new tracking algorithm chooses the

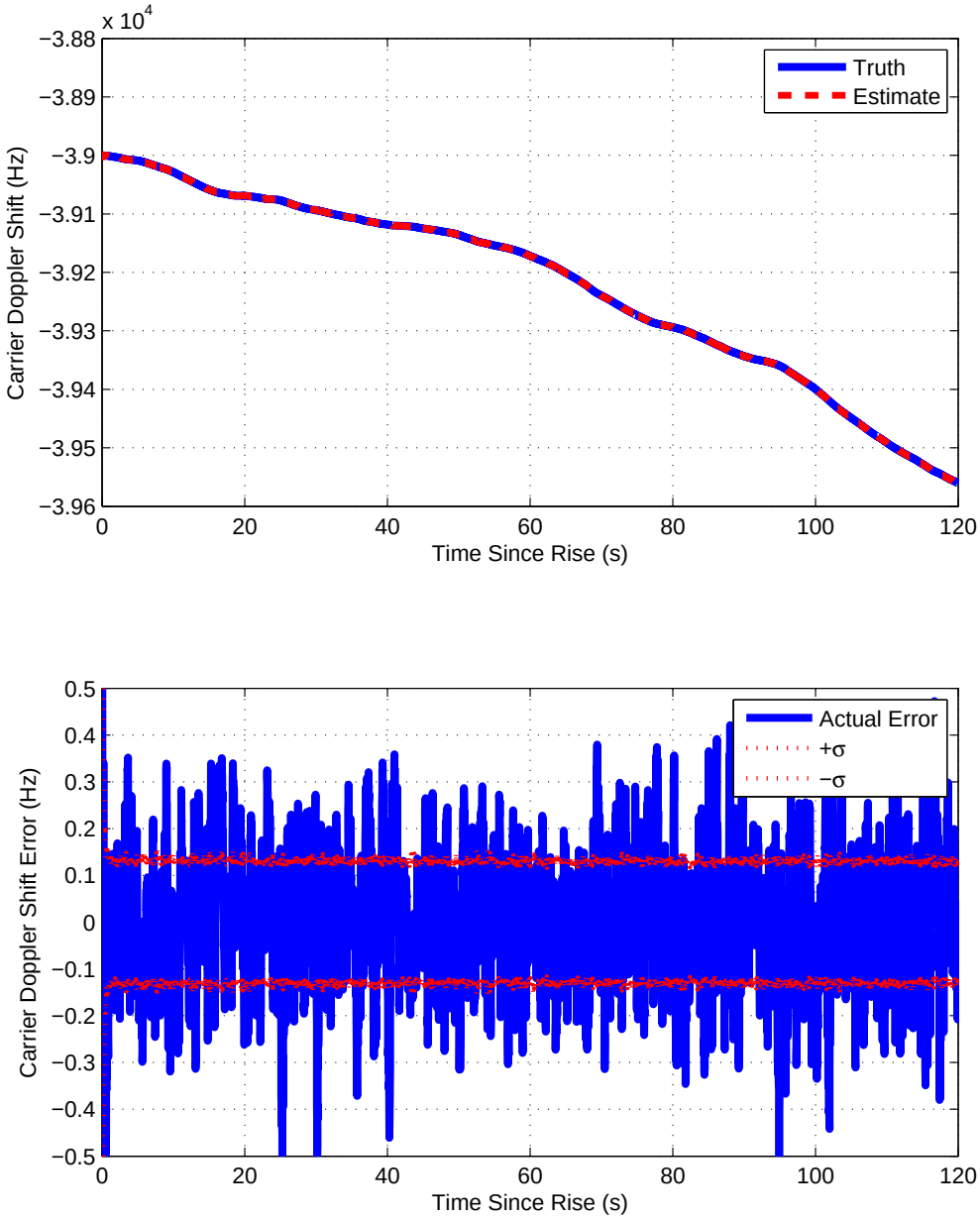

Figure 2.13: Top: Time histories of "truth" and estimated carrier Doppler shift for a weak signal. Bottom: Magnified time history of Doppler shift error and its 1-σ bounds.

correlator with the highest accumulation power from the bank of correlators, along with some of its close neighbours in Doppler-shift and code-phase off-

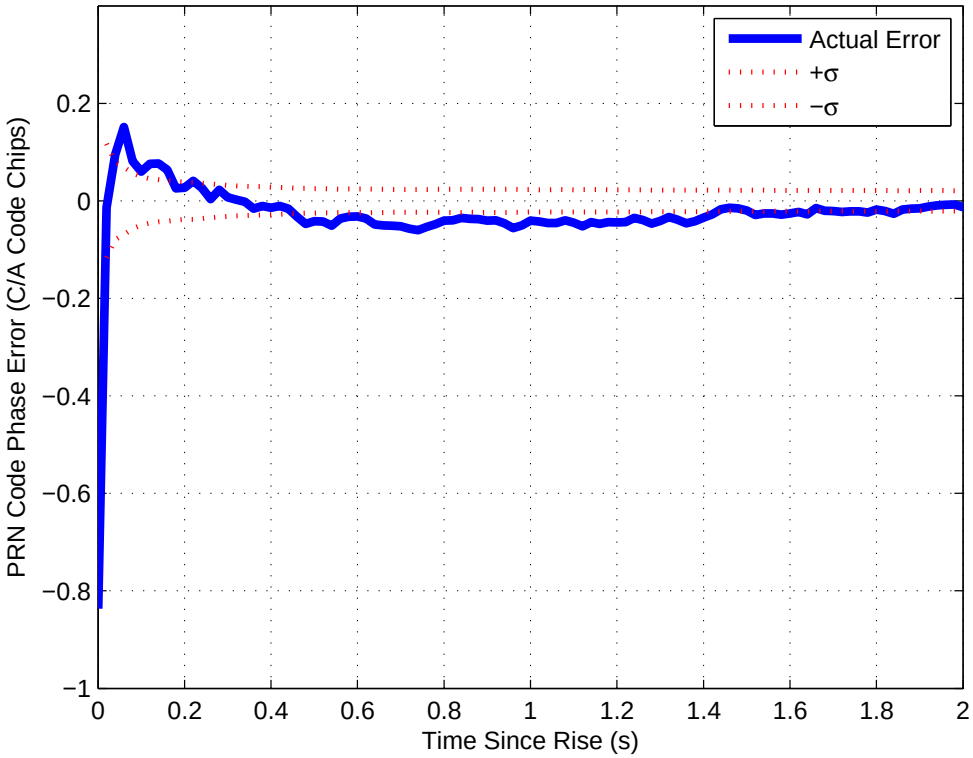

Figure 2.14: Time histories of PRN code phase error and its ±1-σ bounds for a weak signal.

set. The accumulation measurements from these selected correlators are passed through a batch nonlinear optimization algorithm, which is an intermediate estimation step before applying the Kalman filter, and is one that retrieves the most likely signal parameters. The Kalman filter treats the optimized batch states as measurements. It also is used to implement generalized PLL- and DLL-type feedback control laws that provide the necessary inputs for defining the region that must be covered by the bank of correlators.

The algorithm's use of acquisition-like data enables it to achieve robust tracking in the presence of highly dynamic signals with large initial uncertainties. For example, PLL pull-ins from 750 Hz of initial Doppler error can be

achieved using 50 Hz accumulations, as can DLL pull-ins from an initial error of 7 C/A code chips. This capability may be very useful in tracking the signal from a rising GPS satellite received aboard a LEO platform. The resulting limb-scan data, because near-instantaneous tracking can be achieved, will provide information about meteorological conditions within a few meters of the Earth's surface. It may also yield improvements for setting satellites, but improvements would not be as significant, given the ability of standard tracking loops to start with carrier and code lock. Any good PLL/DLL method will have similar steady-state accuracies to this algorithm; the main focus of this chapter is on improving robustness and increasing pull-in ranges.

The algorithm can also track weak signals down to $C/N_0 \cong 25$ dB-Hz, albeit with a bias in the estimated carrier-to-noise ratio. The method needs further refinement, however, if it is to be used with even weaker signals.

Acknowledgment

This work was supported in part by the NASA Office of Space Science under Grant NNX06AC34G. M. Guhathakurta is the grant monitor.

CHAPTER 3

GPS-BASED ATTITUDE DETERMINATION FOR A SPINNING ROCKET

3.1 Abstract

An algorithm is developed for determining the attitude of a spinning sounding rocket. This algorithm is able to track GPS signals with intermittent availability, but nutawith enough accuracy to yield phase observables for the precise, 3-axis attitude determination of a ting rocket. Raw GPS RF front-end data are processed by several filters to accomplish this task. First, a Levenberg-Marquardt algorithm (LMA) estimates GPS observables for multiple satellites by performing a least-squares fit to the accumulation outputs of a bank of correlators. These observables are then used as measurements in a Rauch-Tung-Striebel (RTS) smoother that optimizes estimates of carrier phase, Doppler shift, and code phase. Finally, attitude determination is carried out by another batch filter that uses the single-differenced optimized carrier phase estimates between two antennas and an Euler dynamics model for the torque-free attitude motion of the spinning rocket. This second batch filter implements a combination of a substantially modified form of the LMA and the Least-Squares Ambiguity Decorrelation (LAMBDA) method. This design enables it to deal with integer ambiguities that change over long data gaps between times of carrier phase availability. The algorithm presented in this chapter is applied to recorded RF data from a spinning sounding rocket mission in order to produce attitude quaternion and spin-rate estimates using a pair of antennas separated by a $0.3m$ baseline. These results are confirmed by another set of quaternions and spin-rate vectors independently estimated from magnetometer and horizon crossing indicator data.

Attitude precision on the order of several degrees has been demonstrated.

3.2 Introduction

The work of the present chapter pertains to the task of post-flight attitude determination for a sub-orbital sounding rocket mission: The Magnetosphere-Ionosphere Coupling in the Alfven resonator (MICA) mission. MICA was launched from the Poker Flat Research Range in Fairbanks, Alaska on Feb. 19, 2012. The rocket was spin-stabilized about its minor inertia axis, and it consisted of a main payload and a sub-payload. The sub-payload carried two Micro Pulse 1273FW GPS antennas, linked by about 2 feet of Thermax RG-142S cable to two Zarlink/Plessey GP2015 RF front-ends with a common clock. Raw GPS data from the front-ends were, sequentially, buffered in RAM memory, sent serially to the telemetry encoder, transmitted over an S-band telemetry link, received by the ground station, and recorded in digital recorders. The sub-payload also carried a magnetometer and a horizon crossing indicator (HCI). The magnetometer and HCI can also be used to determine attitude [20], with accuracies demonstrated to be on the order of several degrees, and constitute a way to verify the GPS attitude solution. Figure 3.1 shows the complete rocket system.

The sub-payload ejects from the rest of the rocket at $t \approx 98s$, and its E-field wire booms are fully extended by $t \approx 168s$. This sub-payload is the vehicle for which, at any given epoch time t_0, the attitude quaternion $\boldsymbol{q}_0$ and spin-rate vector $\boldsymbol{\omega}_0$ must be estimated. These quantities at any time t can then be found by propagation of the dynamic model discussed in Section III. However, there are a few major challenges that need to be overcome before this can be achieved

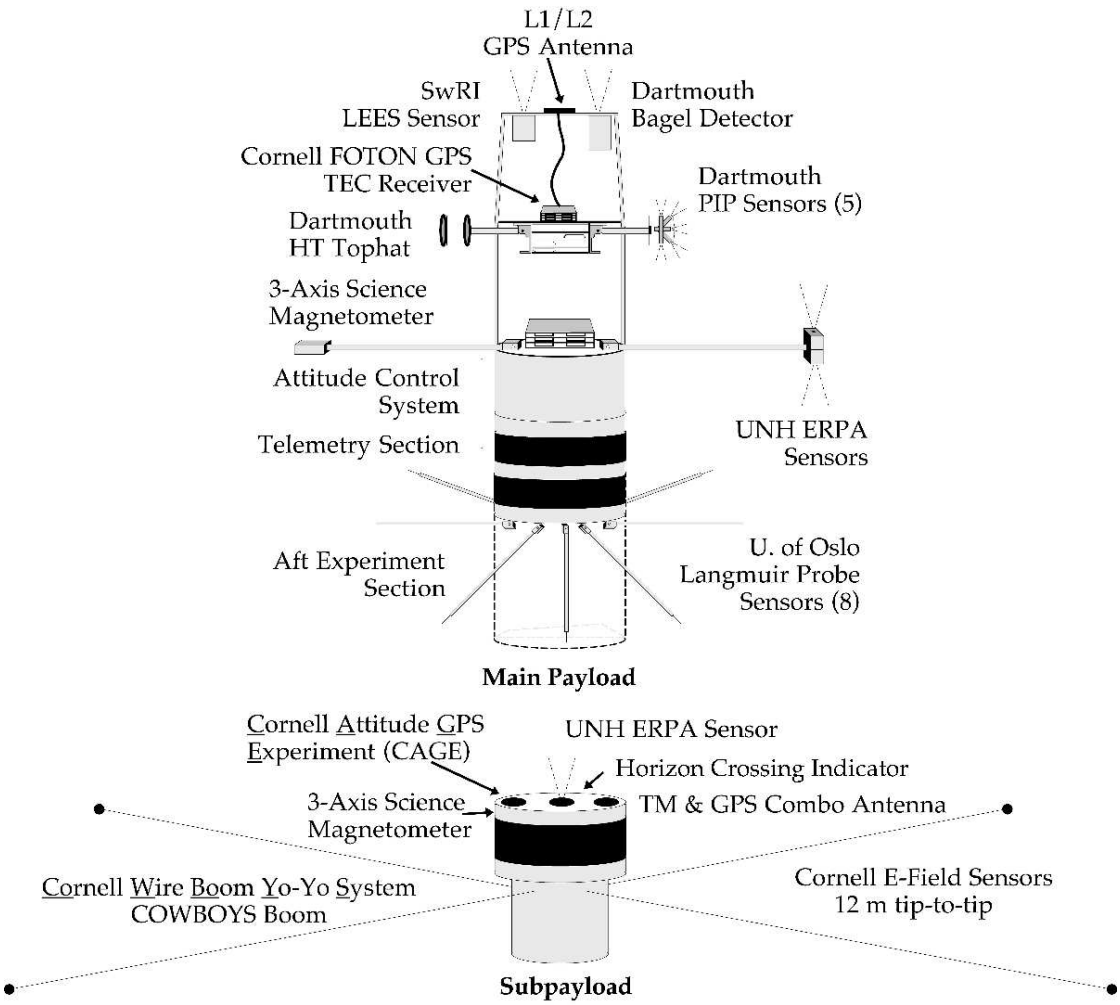

Figure 3.1: MICA sounding rocket layout.

for the MICA mission. First, the rocket undergoes coning motion during its flight. Energy dissipation in the flexible booms leads to growing nutation, a characteristic of minor-axis spinners. Therefore, it cannot be assumed that rotation is restricted to the nominal spin axis. This presents a difficulty because the on-board GPS receiver is connected to only two antennas, so no single epoch can fully determine the required 3-axis attitude and spin-rate. The fact that the rocket is spinning can be exploited to estimate full 3-axis attitude and spin rate from only two antennas by using a nonlinear Euler rotation dynamics model and differential carrier phase, but this leads to further complications.

The second challenge stems from the fact that telemetry bandwidth constraints within the sub-payload system reduce the availability of the received signal data to brief, periodic segments of about $0.0228s$ of data every $0.5s$. This translates into a duty cycle of about 5%. Data gaps in the highly dynamic signal retrieved by a rapidly moving receiver require constant signal re-acquisition. A standard procedure of re-acquisition to initiate a phase-locked tracking loop (PLL) in this situation might need a large amount of processing to acquire with a very fine correlator grid. Alternatively, a PLL might need a large amount of time to settle after a coarse acquisition. The former approach is inefficient and would be impractical for a real-time system, while the latter would not yield useful data from the rocket's short bursts of GPS data availability. These data bursts are also integral to the third and most problematic challenge.

The third challenge lies in the inherent nature of using differential carrier phase measurements: The resolution of integer ambiguities. The considerable number of integer ambiguities in each batch of measurements change every time the signal data stream is interrupted, so they must be re-estimated after every data gap. This amounts to even more integer ambiguities that need to be solved in conjunction with a nonlinear problem.

The main contribution of this chapter is the ability to handle the long data gaps, the nonlinear attitude estimation with nutation dynamics, and the multitude of integer ambiguities, all in one set of algorithms [13]. To the best of the authors' knowledge, the present work is the first to demonstrate two-antenna attitude determination using actual flight data with significant data gaps. Two-antenna attitude determination for spinning spacecraft specifically was first proposed in Ref. [26], based on a fast Fourier transform (FFT) method of es-

timating spin rate and a carrier phase triple-differencing method of estimating orientation. Since then, other works have investigated the use of differential carrier phase in a multi-antenna systems for attitude determination. However, these typically used more than two antennas, needed continuous data from phase-locked loops for long periods of time (and consequently could not deal with data gaps and frequent changes of integer ambiguities) [16], worked only with simulated and ground-experiment-based data, or considered only spacecrafts that rotate much more stably about their major axes (where coning is not significant) [1].

The estimation strategy uses a cascaded approach. This approach starts with specialized GPS signal processing algorithms that are specifically tailored to yield maximum data return from short spans of raw RF front-end outputs, the theory behind which is given in detail in Ref. [12]. It has been shown that this GPS signal processing strategy is able to provide near-instantaneous, accurate tracking, and therefore avoid the loss of information associated with a traditional initial acquisition. It uses correlator banks that maintain enough offsets in Doppler shift and code phase to account for the uncertainty in the location of the signal's correlation peak. Accurate GPS observables at this peak are found by an LMA batch filter and further optimized by an RTS smoother. One of the observables is carrier phase, which is single-differenced between the two antennas to eliminate common-mode receiver errors and provide sub-centimetre relative position accuracy. This is done for multiple satellites to ensure that the resulting attitude determination problem is observable. The raw phase difference measurements between the two antennas for each satellite are compared to a differential carrier phase model in a final LMA nonlinear least-squares batch filter that includes a nonlinear dynamics model to simultaneously estimate $\boldsymbol{q}_0$,

ω_0, and the integer ambiguities.

The remainder of the chapter is divided into four sections. The next section examines the GPS signal processing that extracts the carrier phase measurements via the LMA batch filter and RTS smoother. Section III defines the carrier phase model used by the final attitude determination batch filter. Section IV describes this latter batch filter itself, and Section V presents the results from applying the entire algorithm to the data recorded for MICA.

3.3 GPS Signal Data Processing

3.3.1 Levenberg-Marquardt Batch Optimization to Extract Carrier Phase and Doppler from a Bank of GPS Correlators

As previously mentioned, GPS signal data is only available for segments of approximately $0.0228s$ every $0.5s$. To cope with the data interruptions between these spans, the algorithm begins by performing an FFT acquisition [33] at the beginning of each segment. The Doppler shift output of this acquisition can be a rough estimate, as only the general region containing the correlation peak needs to be located. A correlator bank comprised of discrete points within the Doppler-shift/code-phase space expands about the FFT acquisition's estimated pair of Doppler shift and code phase, with as many offsets in each direction as a 3-σ uncertainty in these two observables. The calculations performed by the correlator bank are comparable to a not-quite brute-force acquisition, and they return in-phase and quadrature accumulations at each Doppler/code-

phase point. The LMA nonlinear batch filter then acts as a multi-correlation vector discriminator of carrier Doppler shift, code phase, carrier phase and carrier amplitude. It finds the best fit of an accumulation measurement model to the discrete correlations in order to yield the four observables. The Doppler shift and code phase determined by this method correspond to the expected point of peak accumulation power. This is illustrated in Fig. 2, where the red dots are the points where the correlator computes accumulations. In effect, the LMA interpolates between them to produce the observables at the theoretical peak. These observables then become the measurements of an extended Kalman filter/smoother that is used to produce tracking-like outputs from a sequence of such measurements, an outline of which follows.

3.3.2 Rauch-Tung-Striebel Smoothing in Lieu of PLL Tracking

A nonlinear extended Kalman filter and the associated Rauch-Tung-Striebel smoother are used to create a smoothed sequence of carrier phase measurements over each $0.0228s$ segment. The theory used to implement the Kalman filter and the RTS smoother are modest adaptations of the square-root information filtering/smoothing techniques of Ref. [5]. The discussion that follows presumes familiarity of those techniques. The nonlinear extended Kalman filter employs the measurement model associated with the outputs of the LMA and a dynamics model for carrier phase, carrier Doppler shift, rate of change of carrier Doppler shift, code phase error, and carrier amplitude [12]. The dynamics model takes the linearized form:

$$\boldsymbol{x}_{k+1} = F_k \boldsymbol{x}_k + \Gamma_k \boldsymbol{v}_k \tag{3.1}$$

where $\boldsymbol{x}_k = (\phi \quad \omega_D \quad \alpha \quad \Delta t_s \quad A)_k^T$ are, respectively, the states beat carrier phase, carrier Doppler shift, rate of change of carrier Doppler shift, code phase error, and carrier amplitude at the start of the k^{th} accumulation interval that spans time $\delta t_k = t_{k+1} - t_k$. $\boldsymbol{v}_k$ is a discrete-time Gaussian process noise vector that models dynamic disturbances. The equations used to derive F_k, the linearized dynamic model state transition matrix, and Γ_k, the dynamic model process noise matrix, are given in Ref. [12]. The Kalman filter's linearized measurement model takes the form:

$$\boldsymbol{y}_k = H_{xk}\boldsymbol{x}_k + H_{vk}\boldsymbol{v}_k + \boldsymbol{w}_k \tag{3.2}$$

where the measurement vector $\boldsymbol{y}_k = (\omega_{Davg} \quad \Delta t_{smid} \quad \phi_{avg} \quad A_{avg})_k^T$ consists of the Doppler shift, code phase error, beat carrier phase, and carrier amplitude, averaged over the k^{th} accumulation interval, which are the optimal outputs of the LMA. The matrices H_{xk} and H_{vk} are given in Ref. [12]. $\boldsymbol{w}_k$ is the discrete-time Gaussian measurement noise vector. The normalized square-root information filter equation uses Eqs. (3.1) and (3.2) to combine the separated dynamic propagation and measurement update steps of Ref. [5] to deal with the presence of process noise $\boldsymbol{v}_k$ in the measurement model of Eq. (3.2). This combined equation is

$$\begin{pmatrix} R_{vvk} & 0 \\ -\hat{R}_{xxk}F_k^{-1}\Gamma_k & -\hat{R}_{xxk}F_k^{-1} \\ R_y^{-T}H_{vk} - R_y^{-T}H_{xk}F_k^{-1}\Gamma_k & R_y^{-T}H_{xk}F_k^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_k \\ \boldsymbol{x}_{k+1} - \bar{\boldsymbol{x}}_{k+1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ R_y^{-T}\boldsymbol{\nu}_k \end{pmatrix} \tag{3.3}$$

where R_{vvk} is the process noise square-root information matrix, $\hat{R}_{xxk}$ is the a posteriori state square-root information matrix, and R_y is the square-root of the measurement error covariance matrix. $\bar{\boldsymbol{x}}_{k+1} = F_k\hat{\boldsymbol{x}}_k$ is the filter's a priori prediction of the state vector at sample $k+1$, $\hat{\boldsymbol{x}}_k$ is its a posteriori estimate at sample k, and the vector innovation is the difference $\boldsymbol{\nu}_k = \boldsymbol{y}_k - H_{xk}\hat{\boldsymbol{x}}_k$. Equation (3.3) is Q-R

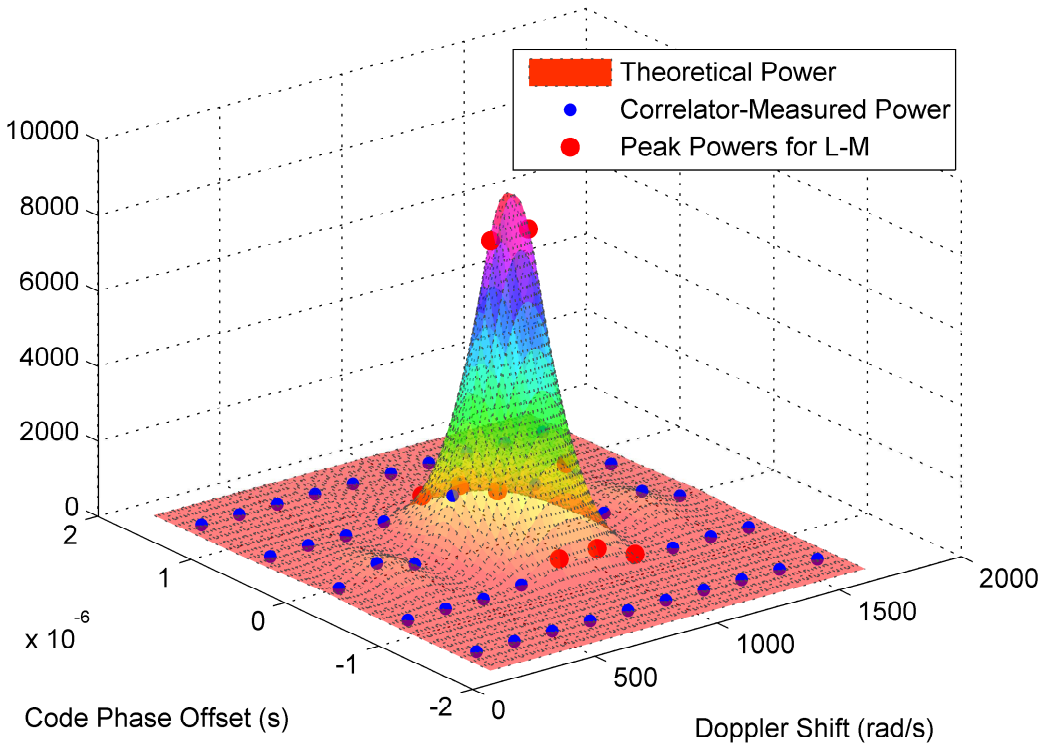

Figure 3.2: Measured power of accumulations from a correlator bank, superimposed on theoretical power.

factorized to yield

$$\begin{pmatrix} \hat{R}_{vvk} & \hat{R}_{vxk} \\ 0 & \hat{R}_{xxk+1} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_k \\ \boldsymbol{x}_{k+1} - \bar{\boldsymbol{x}}_{k+1} \end{pmatrix} = Q^T \begin{pmatrix} 0 \\ 0 \\ R_y^{-T} \boldsymbol{v}_k \end{pmatrix} = \begin{pmatrix} \boldsymbol{\zeta}_{vk} \\ \boldsymbol{\zeta}_{xk+1} \\ \boldsymbol{\zeta}_{rk+1} \end{pmatrix} \tag{3.4}$$

which can be solved to determine the a posteriori estimates $\hat{\boldsymbol{x}}_{k+1} = \hat{R}_{xxk+1}^{-1} \boldsymbol{\zeta}_{xk+1}$ and $\hat{\boldsymbol{v}}_k = \hat{R}_{vvk}^{-1} \left(\boldsymbol{\zeta}_{vk} - \hat{R}_{vxk} \hat{\boldsymbol{x}}_{k+1} \right)$.

Before the filter equations can be implemented correctly, however, there are two issues regarding the phase component of the innovation $\boldsymbol{v}_k$ that needs to be rectified. First, GPS navigation data bits can add half-cycle errors to the phase component of the innovation, so the absolute value of the phase innovation

must be forced to be less than $\pi/2$. Second, the ϕ_{avgk} component of the phase innovation is constrained to lie between $-\pi$ and π by the LMA calculation. This can give rise to aliasing, which must be removed by phase unwrapping. To implement data bit wipe-off and phase unwrapping, the data bit is assumed to be $d_k = 1$ at the beginning of the segment. Then two hypotheses are considered for the phase component of the innovation vector at each successive sample:

$$\nu_{Araw} = \phi_{avgk} - \bar{\phi}_{avgk} + (1 - d_k)\frac{\pi}{2} \tag{3.5}$$

$$\nu_{Braw} = \phi_{avgk} - \bar{\phi}_{avgk} + (1 + d_k)\frac{\pi}{2} \tag{3.6}$$

where $\bar{\phi}_{avgk}$ is the Kalman filter's a priori expected phase determined from its measurement model. Hypothesis A implies that d_{k+1} is the same as d_k, while hypothesis B implies that a data bit flip has occurred. Next, the actual innovation, assuming each hypothesis is correct, is constrained to be less than π:

$$\nu_A = \nu_{Araw} - 2\pi round\left(\frac{\nu_{Araw}}{2\pi}\right) \tag{3.7}$$

$$\nu_B = \nu_{Braw} - 2\pi round\left(\frac{\nu_{Braw}}{2\pi}\right) \tag{3.8}$$

Equations (3.7) and (3.8) constitute a sensible approach to phase unwrapping in the context of this Kalman filter. Each scalar innovation, ν_A and ν_B, is substituted for the phase part of $\boldsymbol{\nu}_k$ in Eq. (3.4) to obtain the residuals $\boldsymbol{\zeta}_{Ark+1}$ and $\boldsymbol{\zeta}_{Brk+1}$.

Finally, if $|\nu_B| < |\nu_A|$ and $\|\boldsymbol{\zeta}_{Ark+1}\|^2 - \|\boldsymbol{\zeta}_{Brk+1}\|^2 > 1$ then a data bit flip is assumed, and the next data bit estimate is $d_{k+1} = -d_k$. The second of the two inequalities checks the sum of the measurement residuals to see if they are significantly different in an ad-hoc way. At times when ν_A and ν_B are close in terms of multiples of $\pi/2$ but there has not been a bit flip, this quantity will be close to zero. If these two conditions are not satisfied, then the next data bit estimate is $d_{k+1} = d_k$.

Although this procedure could allow data bit flips to occur more often than the known $50Hz$ rate, when applied to actual data, the $50Hz$ rate is respected.

The a posteriori Kalman filter states in $\hat{\boldsymbol{x}}_{k+1}$ define the centre of the correlator bank for the interval from t_{k+1} to t_{k+2}, and this whole process is repeated for this next accumulation interval. This process repeats recursively until the end of the $0.0228s$ segment is reached, and constitutes the forward pass of the Kalman-filter/RTS-smoother. As the data segment is being processed, the first row of Eq. (3.4) is retained for each accumulation interval. At the end of the segment, the available information for smoothing consists of the set of first rows from all of the sample intervals and the a posteriori state information at the segment's last sample time. The RTS smoother combines this information with the dynamics model in Eq. (3.1) to produce the smoothed time history $\boldsymbol{x}_k^*$ for each k of the segment. This yields a smoothed time history of the carrier phase, carrier Doppler shift, rate of change of Doppler shift, code phase, and carrier amplitude for each antenna and each satellite. Note the data gaps of the original signal separate the smoothed time histories of individual segments.

At the first segment, the Kalman filter is initialized with states $\hat{\boldsymbol{x}}_0 = (0 \quad \omega_{D0} \quad 0 \quad 0 \quad 0)$ and the associated square-root information matrix

$$\hat{R}_{xx0} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0.005 & 0 & 0 & 0 \\ 0 & 0 & 0.0005 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.9}$$

Only the Doppler shift state begins with a non-zero quantity ω_{D0}, which is the value given by the initial FFT acquisition. Its corresponding element in $\hat{R}_{xx0}$ im-

plies a standard deviation $\sigma \approx 32Hz$. The standard deviation needs to be large enough such that 3-σ encompasses the error allowed by the acquisitions grid resolution in Doppler shift. This error was chosen to be about $75Hz$. Similarly, code phase error, which has a very reasonable initial estimate of 0 $C/A\ Chips$, is covered by $\sigma = 0.2\ C/A\ Chips$. There is no initial information about the rate of change of Doppler shift, but it is assumed to be constant, and its standard deviation of about $318Hz/s$ is more than enough to cover its uncertainty. There is also no initial information about carrier phase and carrier amplitude, but these quantities are solved analytically by the LMA, and can start with infinite uncertainty.

3.4 Differential Carrier Phase Measurement Model

It has been shown that a high-fidelity model for the single-differenced, total carrier phase between two antennas A and B for the same GPS satellite j is [38]

$$\begin{aligned} \lambda\Delta\phi^j_{AB} = &\left(1 + a^j_{f1}\right)\Delta\rho^j_{AB} + c\left(1 + a^j_{f1}\right)(\Delta\delta t_{RAB} + \Delta t_{lnAB}) \\ &- ca^j_{f1}\Delta t_{RAB} + \lambda\Delta N^j_{AB} - \frac{40.3\Delta TEC^j_{AB}}{f^2\left(1 + a^j_{f1}\right)} \\ &- \lambda\left(\Delta\delta\phi^j_{pwuAB} + \Delta\delta\phi^j_{mpAB} + \Delta n^j_{AB}\right) \end{aligned} \tag{3.10}$$

$\Delta\,()_{AB} = ()_A - ()_B$ denotes the difference in the given quantity between antennas A and B. λ is the nominal carrier wavelength, time parameter a^j_{f1} is part of satellite j's navigation message, ρ^j_A is the range from receiver A to satellite j, c is the speed of light, δt_{RA} is receiver A's clock error, t_{lnA} is the line bias, t_{RA} is the receiver time, ΔN^j_{AB} is the carrier phase integer ambiguity between antennas A and B for satellite j, f is the carrier frequency, $\delta\phi^j_{pwuAB}$ is a phase wind up

error due to circular polarization, and the last two terms are the errors due to multipath and thermal noise.

Since the two receivers are linked in this case, the differential receiver clock error and differential receiver clock time drops out. The closeness of the antennas makes the difference in the ionospheric correction $\Delta TEC^j_{AB} = 0$. The antennas are aligned, so the differential phase wind-up term is also zero. The a^j_{f1} parameter is usually negligible over typical baseline lengths between antennas A and B. Multipath can be significant but can be absorbed into the thermal noise term.

All the common-mode biases can be lumped into a real-valued constant $\delta\rho_0 \equiv c\Delta t_{lnAB} + \Delta N^1_{AB}(t_1)$, the last term of which is chosen arbitrarily for some satellite and time, since it cannot be separated observably from the other terms. The differential range model takes the form $\Delta\rho^j_{AB} = \hat{\boldsymbol{\rho}}^{jT} A^T(\boldsymbol{q})\,\boldsymbol{b}_{AB}$, where $\boldsymbol{b}_{AB}$ is the known baseline vector pointing from antenna B to A in spacecraft coordinates, A is the inertial-to-spacecraft direction cosine coordinate transformation matrix, which can be computed from the attitude quaternion $\boldsymbol{q}$ [50], and $\hat{\boldsymbol{\rho}}^j$ is the unit line-of-sight (LOS) vector pointing toward satellite j from the rocket in inertial coordinates. Substituting these quantities into Eq. (3.10) and eliminating negligible terms, the measurement model becomes

$$\lambda\Delta\phi^j_{AB} = \hat{\boldsymbol{\rho}}^{jT} A^T(\boldsymbol{q})\,\boldsymbol{b}_{AB} + \delta\rho_0 + \lambda\delta N^j_{ABp} + w^j \tag{3.11}$$

where the essentially double-differenced $\delta N^j_{ABp} = \Delta N^j_{AB}(p) - \Delta N^1_{AB}(1)$ is assumed to be constant over each data arc $p = 1, ..., P$. Note that the model itself is still in single-differenced form, which is convenient due to the already eliminated receiver clock errors, and the double-differencing in the ambiguities arise from the definition of $\delta\rho_0$. Equation (3.11) is the time-dependent measurement model

with measurement white noise w^j, whose covariance is the sum of the state covariance matrix elements given by the RTS smoothers for receivers A and B. Of course, using the outputs of a smoother somewhat violates the assumption of white noise, but this violation does not appear to cause serious problems in the attitude determination filter. This model's geometries are illustrated in Fig. 3.3. The LOS vector is calculated from a secondary GPS receiver on the sub-payload. The bias $\delta\rho_0$ is assumed to be constant over all data arcs, and the baseline vector $\boldsymbol{b}_{AB}$ is fixed for all times. The main usefulness of Eq. (3.11) is its dependence on the attitude quaternion $\boldsymbol{q}$ in its first term. The unknown quantities $\delta\rho_0$ and δN^j_{ABp} are nuisance parameters that will have to be estimated as part of the attitude solution. One of the challenges of the present effort is the fact that the nuisance double-differenced integer ambiguities must be re-estimated for each independent data segment $p = 1, ..., P$.

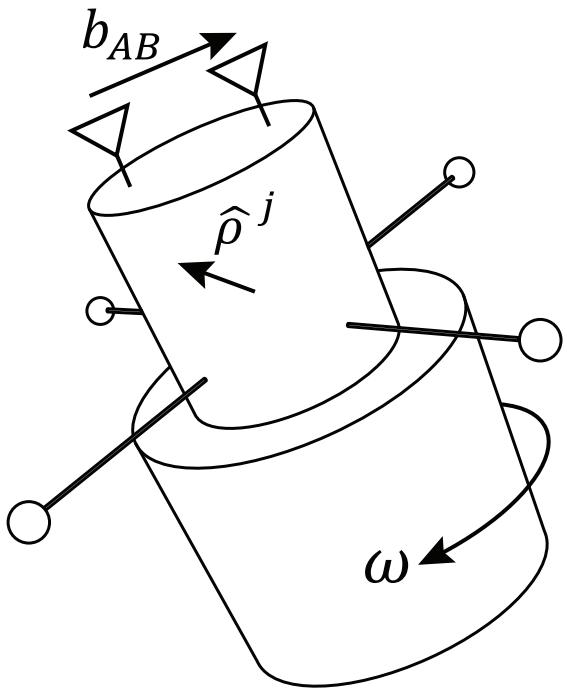

Figure 3.3: Diagram of antennas, baseline vector, and LOS unit vector of spinning sub-payload.

Figure 3.4 shows a set of segments of differential carrier phase measurements that are output by the correlator-bank/LMA/Kalman-filter/RTS-smoother calculations. Each column of points is a separate $0.0228s$ segment that looks like Fig. 3.5 if enlarged. This time history of measurements is compared to the measurement model of Eq. (3.11) in the batch filter of the next section. The batch filter will determine the attitude time history that best fits these measurements. In this raw state, however, it is not at all obvious that these correspond to any sensible attitude motions.

In order to relate the unknown quaternion $\boldsymbol{q}_0$ and the unknown angular velocity $\boldsymbol{\omega}_0$, at an epoch time t_0, to the unknown quaternion in Eq. (3.11) at the corresponding single-differenced carrier phase measurement time, quaternion and Euler dynamics equations are used:

$$\dot{\boldsymbol{q}}(t) = \frac{1}{2}\begin{bmatrix} [\boldsymbol{\omega}(t)]_\times^T & \boldsymbol{\omega}(t) \\ \boldsymbol{\omega}^T(t) & 0 \end{bmatrix}\boldsymbol{q}(t) \tag{3.12}$$

$$\dot{\boldsymbol{\omega}}(t) = \left[I_{mom}^{-1}\,[\boldsymbol{\omega}(t)]_\times\right]\boldsymbol{\omega}(t) \tag{3.13}$$

where $[\boldsymbol{\omega}(t)]_\times$ is the skew-symmetric cross-product matrix associated with the 3-dimensional rate vector $\boldsymbol{\omega}(t)$, and I_{mom} is the 3×3 symmetric positive definite spacecraft moment of inertia matrix. Equations (3.12) and (3.13) are numerically integrated together to yield the following equation for the quaternion at a general time:

$$\boldsymbol{q}(t; \boldsymbol{q}_0, \boldsymbol{\omega}_0) = \Phi_{qq}(t, \boldsymbol{\omega}_0, t_0)\boldsymbol{q}_0 \tag{3.14}$$

Note that this equation is linear in $\boldsymbol{q}_0$ but nonlinear in $\boldsymbol{\omega}_0$. Substitution of Eq. (3.14) into Eq. (3.11) gives the direct dependence of differential carrier phase

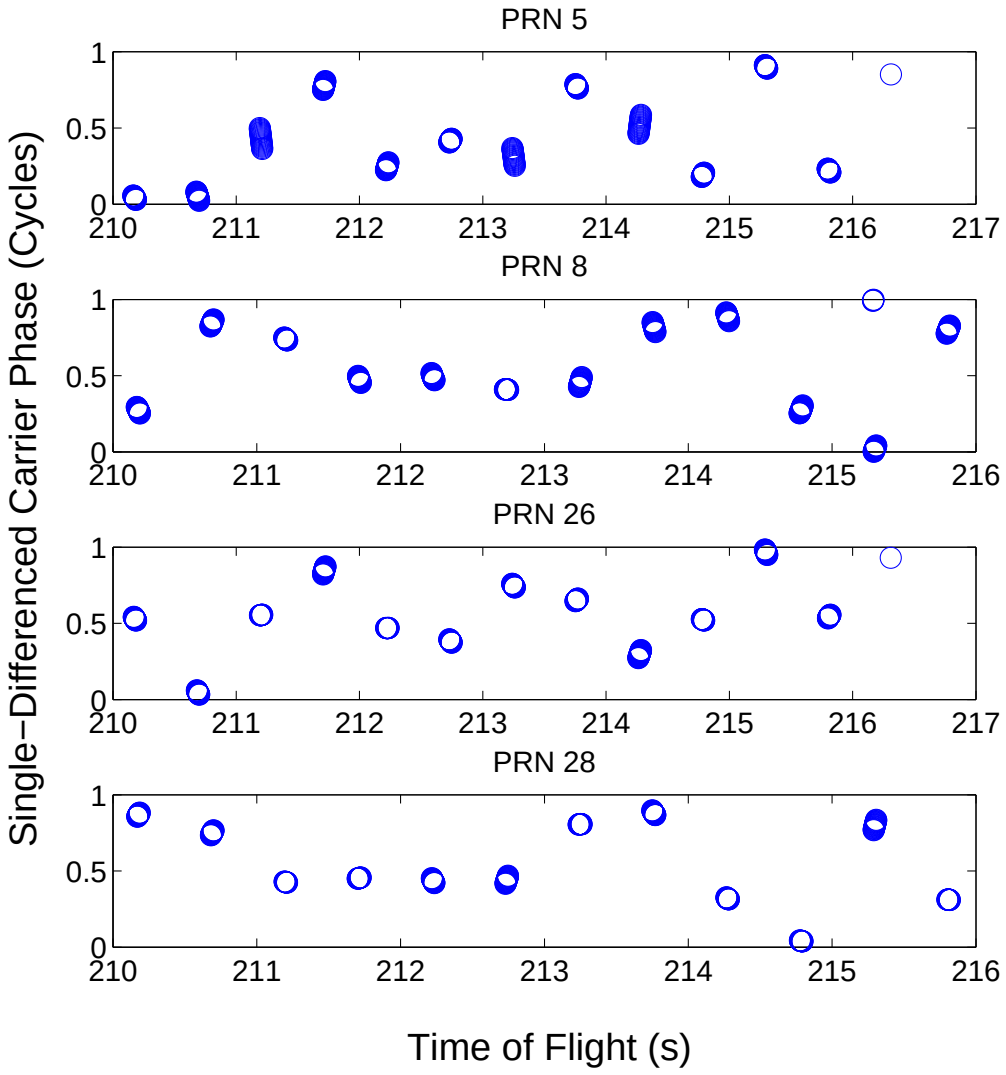

Figure 3.4: A set of $0.0228s$ segments of single-differenced RTS smoother carrier phases that constitute the inputs to an attitude batch filter.

on $\boldsymbol{q}_0$ and $\boldsymbol{\omega}_0$. It is nonlinear in $\boldsymbol{q}_0$ because of nonlinearities in the $A(\boldsymbol{q})$ formula, and it is nonlinear in $\boldsymbol{\omega}_0$ due to the nonlinear effects inherent in integrating Eqs. (3.12) and (3.13), and due to the nonlinearities in the $A(\boldsymbol{q})$ formula.

3.5 Attitude Batch Filter

The development of the attitude determination batch filter begins by stacking all the single-differenced measurements into a vector

$$rClz = \lambda \begin{bmatrix} \Delta\phi_{AB}^{1}(t_1^1) \\ \Delta\phi_{AB}^{1}(t_2^1) \\ \vdots \\ \Delta\phi_{AB}^{1}(t_{K_1}^1) \\ \Delta\phi_{AB}^{2}(t_1^2) \\ \vdots \\ \Delta\phi_{AB}^{M}(t_{K_M}^M) \end{bmatrix} \tag{3.15}$$

for all satellites $j = 1, ..., M$ and all sample times. Note, $t_1^j, ..., t_{K_j}^j$ are the K_j sample times at which single-differenced data are available for satellite j. The nonlinear part of Eq. (3.11) is isolated as another stacked vector

$$\boldsymbol{h}(\boldsymbol{q}_0, \boldsymbol{\omega}_0) = \begin{bmatrix} \hat{\boldsymbol{\rho}}^{1T}(t_1^1) A^T \left\{ \boldsymbol{q}(t_1^1; \boldsymbol{q}_0, \boldsymbol{\omega}_0) \right\} \boldsymbol{b}_{AB} \\ \hat{\boldsymbol{\rho}}^{1T}(t_2^1) A^T \left\{ \boldsymbol{q}(t_2^1; \boldsymbol{q}_0, \boldsymbol{\omega}_0) \right\} \boldsymbol{b}_{AB} \\ \vdots \\ \hat{\boldsymbol{\rho}}^{1T}(t_{K_1}^1) A^T \left\{ \boldsymbol{q}(t_{K_1}^1; \boldsymbol{q}_0, \boldsymbol{\omega}_0) \right\} \boldsymbol{b}_{AB} \\ \hat{\boldsymbol{\rho}}^{2T}(t_1^2) A^T \left\{ \boldsymbol{q}(t_1^2; \boldsymbol{q}_0, \boldsymbol{\omega}_0) \right\} \boldsymbol{b}_{AB} \\ \vdots \\ \hat{\boldsymbol{\rho}}^{MT}(t_{K_M}^M) A^T \left\{ \boldsymbol{q}(t_{K_M}^M; \boldsymbol{q}_0, \boldsymbol{\omega}_0) \right\} \boldsymbol{b}_{AB} \end{bmatrix} \tag{3.16}$$

Given these $\boldsymbol{z}$ and $\boldsymbol{h}$ definitions, and given the square-root information matrix for the measurement noise in $\boldsymbol{z}$, R_a, one can define the following batch

weighted least-squares cost function

$$
\begin{aligned}
J(\boldsymbol{q}_0, \boldsymbol{\omega}_0, \delta\rho_0, \delta\boldsymbol{N}) = \\
&\frac{1}{2}\left\{R_a\left[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{q}_0, \boldsymbol{\omega}_0) - A_{\delta\rho}\delta\rho_0\right.\right. \\
&\left.\left. - A_n\delta\boldsymbol{N}\right]\right\}^T\left\{R_a\left[\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{q}_0, \boldsymbol{\omega}_0) - A_{\delta\rho}\delta\rho_0 - A_n\delta\boldsymbol{N}\right]\right\}
\end{aligned}
\tag{3.17}
$$

where $\delta\boldsymbol{N} = \left[\delta N^2_{AB1} \quad \delta N^3_{AB1} \quad \ldots \quad \delta N^M_{AB1} \quad \delta N^1_{AB2} \quad \delta N^2_{AB2} \quad \ldots \quad \delta N^M_{ABP}\right]^T$ is the $(P \times M - 1)$-dimensional vector of all double-differenced integer ambiguities in the entire batch, and where $A_{\delta\rho}$ and A_n are coefficient matrices for the constant bias and differential integer ambiguities derived from the measurement model. Note that R_a is the square-root of the inverse of the covariance matrix for the measurement error in $\boldsymbol{z}$.

The batch filter estimates $\boldsymbol{q}_0$ and $\boldsymbol{\omega}_0$ by minimizing the cost in Eq. (3.17) with respect to $\boldsymbol{q}_0$, $\boldsymbol{\omega}_0$, and the nuisance parameters $\delta\rho_0$ and $\delta\boldsymbol{N}$. This is a mixed real/integer nonlinear least-squares problem. The integers, however, enter it linearly. Therefore, a specifically tailored LMA has been developed to solve it, one that is similar to the method used in Ref. [34]. It combines integer linear least-squares methods [45] [37] with Levenberg-Marquardt calculations.

The adapted LMA works with current guesses of $\boldsymbol{q}_0$ and $\boldsymbol{\omega}_0$. Call them $\hat{\boldsymbol{q}}_0$ and $\hat{\boldsymbol{\omega}}_0$. It expresses the true values as sums of current guesses plus perturbations:

$$\boldsymbol{q}_0 = \hat{\boldsymbol{q}}_0 + \Delta\boldsymbol{q}_0 \tag{3.18}$$

$$\boldsymbol{\omega}_0 = \hat{\boldsymbol{\omega}}_0 + \Delta\boldsymbol{\omega}_0 \tag{3.19}$$

$\boldsymbol{h}$ is then expanded in a first-order Taylor series to yield a version of the equa-

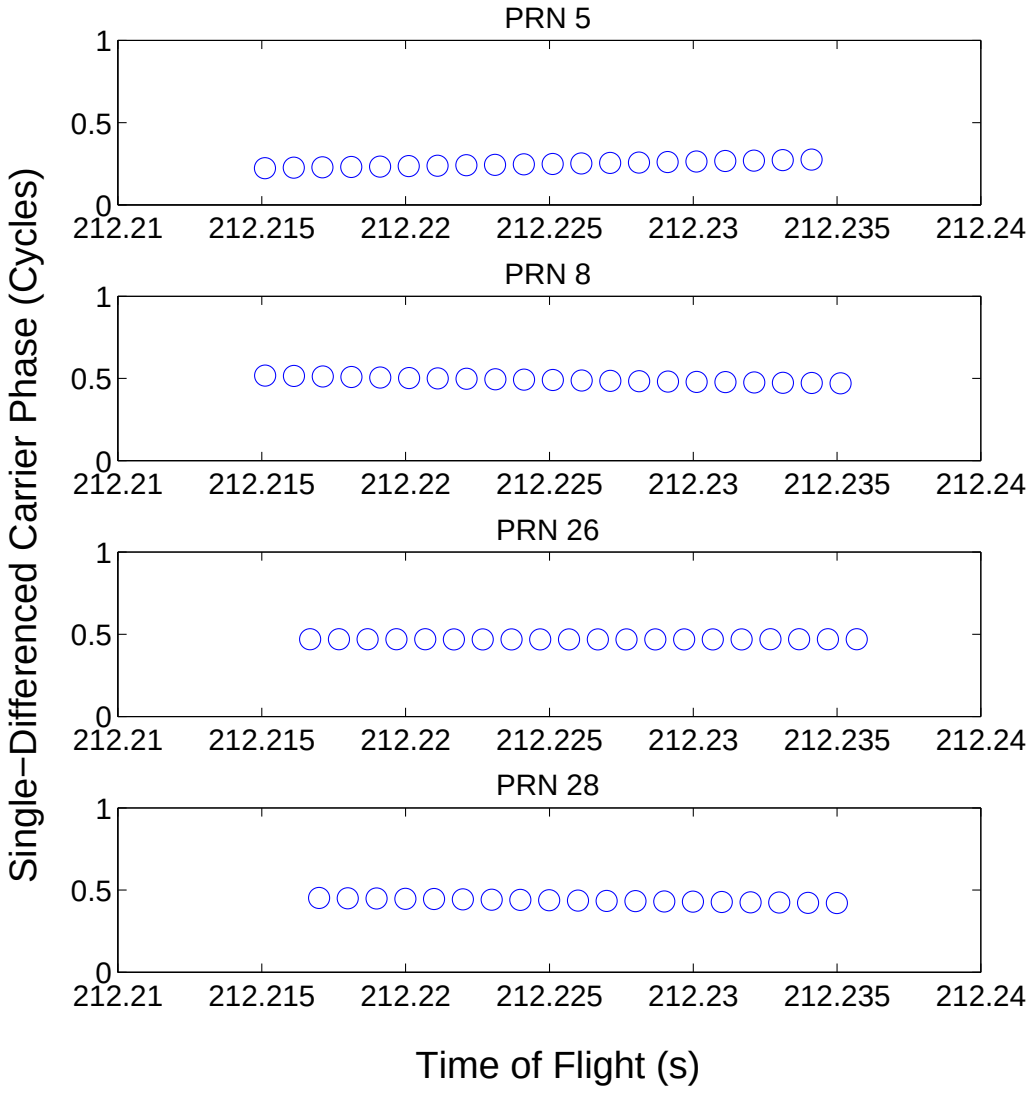

Figure 3.5: Zoomed-in version of Fig. 3.4, showing one $0.0228s$ segment of single-differenced carrier phase measurements.

tions whose squared errors constitute the batch cost function:

$$R_a\left[\boldsymbol{z} - \boldsymbol{h}(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0)\right] = R_a \begin{pmatrix} A_q & A_\omega & A_{\delta\rho} & A_n \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{q}_0 \\ \Delta\boldsymbol{\omega}_0 \\ \delta\rho_0 \\ \delta\boldsymbol{N} \end{pmatrix} \tag{3.20}$$

where A_q and A_ω are the Jacobian matrices of the Taylor-series expansion eval-

uated at $\hat{\boldsymbol{q}}_0$ and $\hat{\boldsymbol{\omega}}_0$. In order to limit the size of the $\Delta\boldsymbol{q}_0$ and $\Delta\boldsymbol{\omega}_0$ perturbations, thereby limiting how far the linearized model deviates from the original, non-linear model, two more rows are added to Eq. (3.20). At the same time, a transformed version of $\Delta\boldsymbol{q}_0$ is used, one that allows implicit enforcement of the usual quaternion normalization constraint. The resulting new equations are

$$\begin{pmatrix} R_a\left[\boldsymbol{z} - \boldsymbol{h}(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0)\right] \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} R_a A_q Q_s & R_a A_\omega & R_a A_{\delta\rho} & R_a A_n \\ \sqrt{\gamma} I_3 & 0 & 0 & 0 \\ 0 & T\sqrt{\gamma} I_3 & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{q}_{0s} \\ \Delta\boldsymbol{\omega}_0 \\ \delta\rho_0 \\ \delta\boldsymbol{N} \end{pmatrix} \tag{3.21}$$

The new last two rows utilize a tuning parameter γ and nominal period of rotation T to enforce soft constraints on the total $\Delta\boldsymbol{q}_{0s}$ and $\Delta\boldsymbol{\omega}_0$ perturbations. Sensible limiting of $\Delta\boldsymbol{q}_{0s}$ and $\Delta\boldsymbol{\omega}_0$ through a wise choice of γ can guarantee convergence for the LMA to a local minimum.

Note that the final linearized model of Eq. (3.21) uses the transformation $\Delta\boldsymbol{q}_0 = (\Delta\boldsymbol{q}_{0s} \quad 0)^T \otimes \hat{\boldsymbol{q}}_0 = Q_s(\hat{\boldsymbol{q}}_0)\Delta\boldsymbol{q}_{0s}$ [35] to eliminate the 4×1 quaternion perturbation $\Delta\boldsymbol{q}_0$ in favour of the 3×1 multiplicative quaternion perturbation $\Delta\boldsymbol{q}_{0s}$, where $\otimes$ denotes quaternion multiplication, and $Q_s(\hat{\boldsymbol{q}}_0)$ is a 4×3 matrix that defines the tangent space to the unit-normalized hypersphere at the point $\hat{\boldsymbol{q}}_0$. This substitution ensures that the perturbations are orthogonal to the nominal quaternion, which guarantees the unit-normalization of $\hat{\boldsymbol{q}}_0 + \Delta\boldsymbol{q}_{0s}$ to first-order in $\Delta\boldsymbol{q}_0$.

The adapted LMA is implemented as follows. First, an initial guess of the pair $\hat{\boldsymbol{q}}_0$ and $\hat{\boldsymbol{\omega}}_0$ is chosen by drawing $\hat{\boldsymbol{q}}_0$ randomly from the space of all unit-normalized 4-vectors, and letting $\hat{\boldsymbol{\omega}}_0$ be the approximate, nominal angular velocity set by spin-stabilization. This is part of a method that guarantees global

convergence. If the initial guesses of $\hat{\boldsymbol{q}}_0$ and $\hat{\boldsymbol{\omega}}_0$ are too far from their true values, the algorithm fails to converge. In order to compensate for this possibility, many initial guesses are tried, and the algorithm is iterated to a local minimum or until it diverges. After many such trials, the minimum that has the lowest overall cost is chosen as the optimal estimate, as an added precaution that can save computation time for very bad guesses.

The algorithm begins by evaluating Eq. (3.21) with the initial guesses of $\hat{\boldsymbol{q}}_0$ and $\hat{\boldsymbol{\omega}}_0$ and $\gamma = 0$ to produce the optimal least-squares perturbations $\Delta\boldsymbol{q}_{0s}$ and $T\Delta\boldsymbol{\omega}_0$. If the total norm of these perturbations is greater than 1, a new pair of initial guesses is chosen as before. This is an ad-hoc test that saves processing time, since large perturbations suggest the algorithm will not converge. If the total norm of the perturbations is less than 1, then the algorithm proceeds to a regular set of iterations.

Given reasonable first guesses of $\hat{\boldsymbol{q}}_0$ and $\hat{\boldsymbol{\omega}}_0$, the modified LMA is applied to iteratively improve them in a manner that decreases the cost function in Eq. (3.17). In order to do this, the algorithm needs a method to evaluate the cost purely as a function of $\hat{\boldsymbol{q}}_0$ and $\hat{\boldsymbol{\omega}}_0$. That is, it needs a method of eliminating the dependence on $\delta\rho_0$ and $\delta\boldsymbol{N}$. This is accomplished by optimizing $\delta\rho_0$ and $\delta\boldsymbol{N}$ for given fixed values of $\hat{\boldsymbol{q}}_0$ and $\hat{\boldsymbol{\omega}}_0$. The optimal values $\delta\rho_{0opt}(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0)$ and $\delta\boldsymbol{N}_{opt}(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0)$ are determined by minimizing the sum of the squared errors in Eq. (3.20), subject to the constraints $\Delta\boldsymbol{q}_0 = 0$ and $\Delta\boldsymbol{\omega}_0 = 0$. This is a mixed real/integer linear least-squares problem that can be solved using the methods in Ref. [37]. The resulting cost function for $\hat{\boldsymbol{q}}_0$ and $\hat{\boldsymbol{\omega}}_0$ is

$$\tilde{J}(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0) = J\left[\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0, \delta\rho_{0opt}(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0), \delta\boldsymbol{N}_{opt}(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0)\right] \tag{3.22}$$

Each iteration of the modified LMA chooses its perturbations $\Delta\boldsymbol{q}_{0s}$ and $\Delta\boldsymbol{\omega}_0$ by

solving for the values of $\Delta\boldsymbol{q}_{0s}$, $\Delta\boldsymbol{\omega}_0$, $\delta\rho_0$, and $\delta\boldsymbol{N}$ that minimize the square of the errors in Eq. (3.21). This too is a mixed real/integer linear least-squares problem that is solvable by the techniques of Ref. [37]. The new attitude and rate guesses are

$$\hat{\boldsymbol{q}}_{0new} = \frac{\hat{\boldsymbol{q}}_0 + Q_s(\hat{\boldsymbol{q}}_0)\Delta\boldsymbol{q}_{0s}}{\|\hat{\boldsymbol{q}}_0 + Q_s(\hat{\boldsymbol{q}}_0)\Delta\boldsymbol{q}_{0s}\|} \tag{3.23}$$

$$\hat{\boldsymbol{\omega}}_{0new} = \hat{\boldsymbol{\omega}}_0 + \Delta\boldsymbol{\omega}_0 \tag{3.24}$$

These guesses are not accepted, however, without first verifying that $\tilde{J}(\hat{\boldsymbol{q}}_{0new}, \hat{\boldsymbol{\omega}}_{0new}) < \tilde{J}(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0)$. If this inequality does not hold true, then the Levenberg-Marquardt parameter γ is increased, and the mixed real/integer least-squares problem associated with Eq. (3.21) is re-solved to generate revised optimal values of $\Delta\boldsymbol{q}_{0s}$ and $\Delta\boldsymbol{\omega}_0$, and revised new guesses $\hat{\boldsymbol{q}}_{0new}$ and $\hat{\boldsymbol{\omega}}_{0new}$.

It can be shown theoretically that $\tilde{J}(\hat{\boldsymbol{q}}_{0new}, \hat{\boldsymbol{\omega}}_{0new})$ will be less than $\tilde{J}(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0)$ for some sufficiently large value of γ. It can also be shown theoretically that the LMA will converge most rapidly near a local minimum in $(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0)$-space if $\gamma = 0$. Therefore, the inner-loop γ adjustment of the LMA starts with $\gamma = 0$ and iteratively increases it, as needed in order to achieve $\tilde{J}(\hat{\boldsymbol{q}}_{0new}, \hat{\boldsymbol{\omega}}_{0new}) < \tilde{J}(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0)$.

The main iterations of the LMA are those that start with given guesses $\hat{\boldsymbol{q}}_0$ and $\hat{\boldsymbol{\omega}}_0$, and that end with the new guesses $\hat{\boldsymbol{q}}_{0new}$ and $\hat{\boldsymbol{\omega}}_{0new}$ after possible iteration to determine γ that forces a cost decrease. At the end of each main iteration, $\hat{\boldsymbol{q}}_{0new}$ and $\hat{\boldsymbol{\omega}}_{0new}$ replace $\hat{\boldsymbol{q}}_0$ and $\hat{\boldsymbol{\omega}}_0$ in preparation for the next iteration. These iterations repeat until the achievable decrement in $\tilde{J}$ becomes negligible. This usually occurs during terminal iterations that use $\gamma = 0$.

Note, each main iteration involves the solution of two related, but slightly different mixed real/integer linear least-squares problems. One solves for $\Delta\boldsymbol{q}_{0s}$,

$\Delta\boldsymbol{\omega}_0$, $\delta\rho_0$, and $\delta\boldsymbol{N}$ using Eq. (3.21). This solution is used to compute the Levenberg-Marquardt increments $\Delta\boldsymbol{q}_0$ and $\Delta\boldsymbol{\omega}_0$. The other solves for $\delta\rho_0$ and $\delta\boldsymbol{N}$ using Eq. (3.20) with its left-hand-side evaluated at $\hat{\boldsymbol{q}}_{0new}$ and $\hat{\boldsymbol{\omega}}_{0new}$, and with $\Delta\boldsymbol{q}_0$ and $\Delta\boldsymbol{\omega}_0$ constrained to be zero. This solution is needed to evaluate $\tilde{J}(\hat{\boldsymbol{q}}_{0new}, \hat{\boldsymbol{\omega}}_{0new})$ in order to determine whether to accept these new solution guesses, or whether to return to Eq. (3.21) using a larger value of γ.

The algorithm usually passes its termination tests after less than ten iterations. Moreover, given that the rocket spin rate is $0.44Hz$, the half-second data gaps can be tolerated without aliasing.

3.6 Attitude Results

The following are the results of applying the attitude batch filter to the carrier phase measurements extracted by the RTS smoother from the raw GPS signal data received aboard MICA. The attitude batch filter finds the best fit of these measurements to the aforementioned carrier phase model by adjusting attitude parameters. Figure 3.6 shows the modeled phases computed from the final solution parameters, and their close fit to the RTS phase data, thus confirming that these data correspond to sensible rigid body attitude motions. The attitude solutions themselves have been compared to the results from a magnetometer and an HCI on the same payload. The data from these two instruments are combined in a separate extended Kalman filter/smoother to yield angular momentum and attitude parameters that are converted into quaternion solutions [20]. Figure 3.7 plots each of the quaternion components from the present filter and from the magnetometer/HCI filter/smoother. The batch filter processes $6s$

data batches once every 20 s, and each batch solution is interpolated over its data batch and extrapolated $14s$ past the end of its batch. The batch solutions are interpolated and extrapolated using the rotational dynamics model in Eqs. (3.12) and (3.13). It would be possible to use more data to implement a contiguous set of batch solutions, but the extra effort seems unnecessary given the good results achieved with this method. The data batch length encompasses at least two periods of rotation, but it is small compared to the time scale of coning angle variations. Each batch solution $(\hat{\boldsymbol{q}}_0, \hat{\boldsymbol{\omega}}_0)$ falls at the beginning of the batch, and is denoted by '$\times$' on Fig. 3.7. The interpolated/extrapolated solutions match closely with those computed from magnetometer/HCI data for the whole flight, although only a portion of the flight is shown graphically. The discrepancies between the attitude results from the GPS and those from the magnetometer/HCI are given in Figs. 3.8, 3.9, and 3.10. Most of the disagreement lies along the spin axis, which may be an artifact of an inaccurate Earth magnetic field model. Even so, the discrepancies are relatively small, with the total angular difference being 4.86°, and the maximum difference being 11.13°. Given that the magnetometer/HCI filter/smoother has an expected accuracy on the order of several degrees, this is an encouraging level of agreement.

Angular velocity plots are presented in Fig. 3.11 for the same two filters, with the ± 1-σ values from the batch filter in red. The upper panel shows the transverse angular velocity oscillations that are characteristic of nutation, and the lower panel shows the angular velocity about the nominal spin axis. Noting the small scale of the vertical axes, the discrepancies are small. The agreement of oscillations in the transverse axes suggests that the batch filter is estimating a reasonable coning angle. In particular, the coning phases agree well and the amplitudes somewhat less well. Given the small amplitudes, this is not a cause

for concern.

Figure 3.12 plots the batch filter's theoretical estimates of its attitude accuracy. That is, it plots the standard deviations of the elements of $\boldsymbol{q}(t; \hat{\boldsymbol{q}}_0), \hat{\omega}_0$. These are inferred from the inverse of the Fisher information matrix for its $\hat{\boldsymbol{q}}_0$ and $\hat{\omega}_0$ estimates. These accuracies correspond to per-axis attitude 1-σ accuracies on the order of 0.22° or less. These are likely optimistic, but their low values confirm that the underlying differential carrier phase measurements and attitude dynamics model cause $\hat{\boldsymbol{q}}_0$ and $\hat{\omega}_0$ to be observable.

3.7 Conclusion

This chapter has presented a way to resolve 3-axis attitude and attitude-rate using only two GPS antennas, despite a highly dynamic environment and a low duty cycle in signal reception – one $0.0228s$ segment of GPS data every $0.5s$ of rocket flight. The 3-axis attitude and rate estimation algorithm solves a mixed real/integer problem, and eliminates the need to first remove the ambiguities at the possible cost of discarding useful information. The efficacy of this chapter's algorithm is demonstrated by application to a set of data from a sounding rocket experiment that includes coning motion. Despite its long signal gaps, the raw GPS data from this mission is successfully tracked by a using a bank of correlators, a Levenberg-Marquardt vector discriminator, and a Rauch-Tung-Striebel smoother to produce accurate differential carrier phases. These phases are processed in a batch filter that solves for the real-valued attitude solution and carrier phase integer ambiguities simultaneously. The results were reasonable when compared with those from a magnetometer and horizon crossing in-

dicator. The RMS attitude difference between the two methods over the whole flight was 4.86°, and the maximum difference was 11.13°.

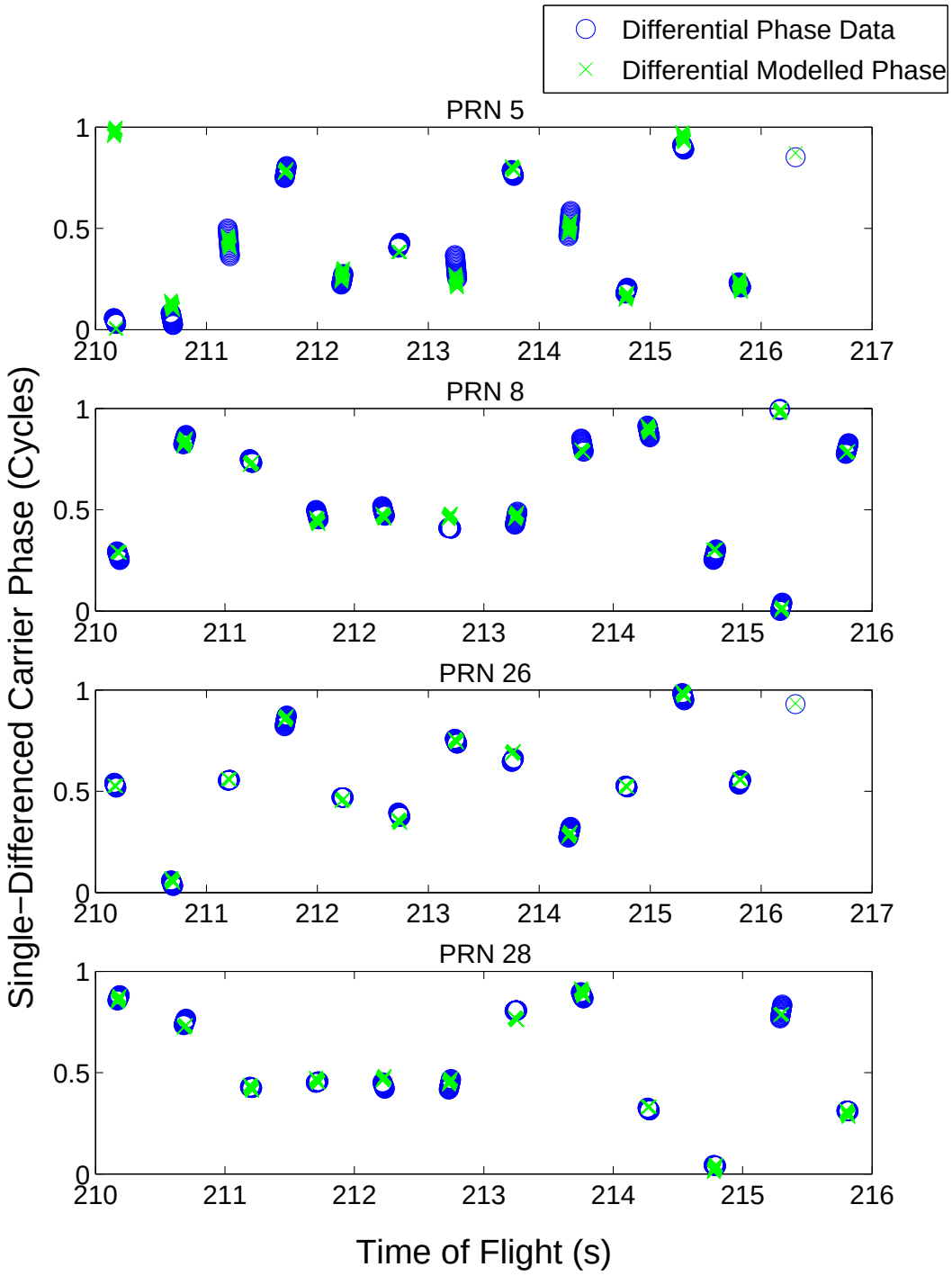

Figure 3.6: Differential carrier phase data from Fig. 3.4 are fit to attitude batch filter's phase model.

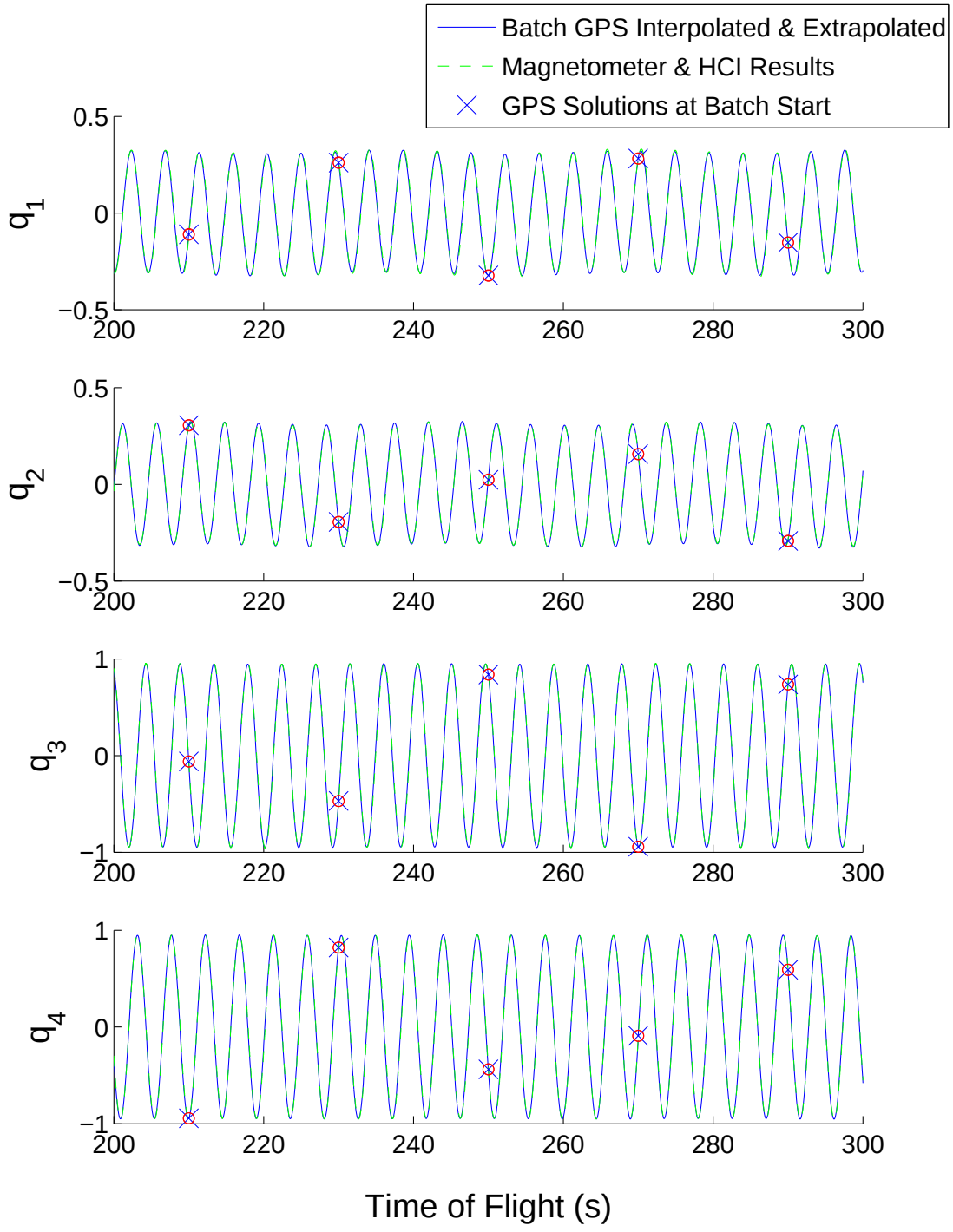

Figure 3.7: Estimated quaternion time histories from the GPS batch filter and from the magnetometer/HCI for comparison.

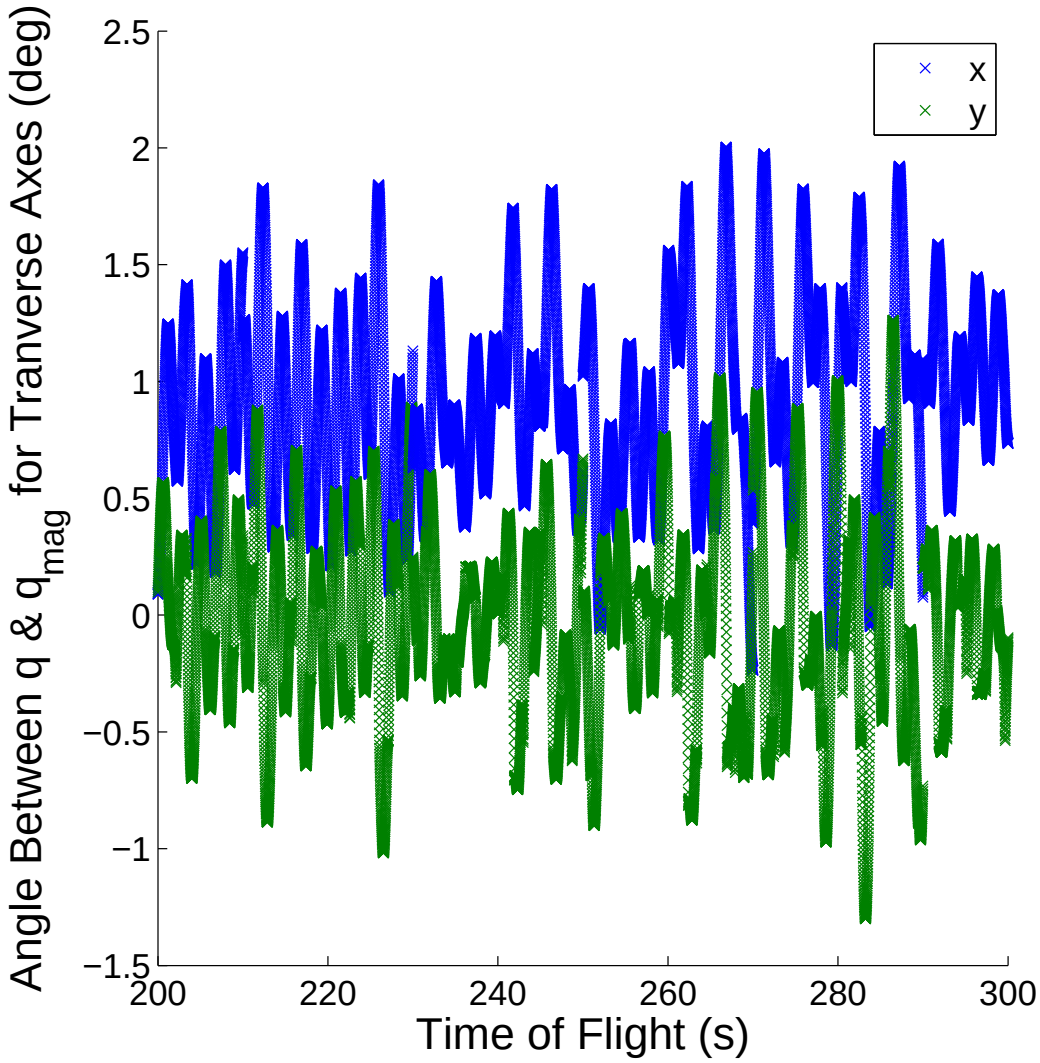

Figure 3.8: Angle discrepancy time history between GPS estimates and magnetometer/HCI estimates for transverse axes.

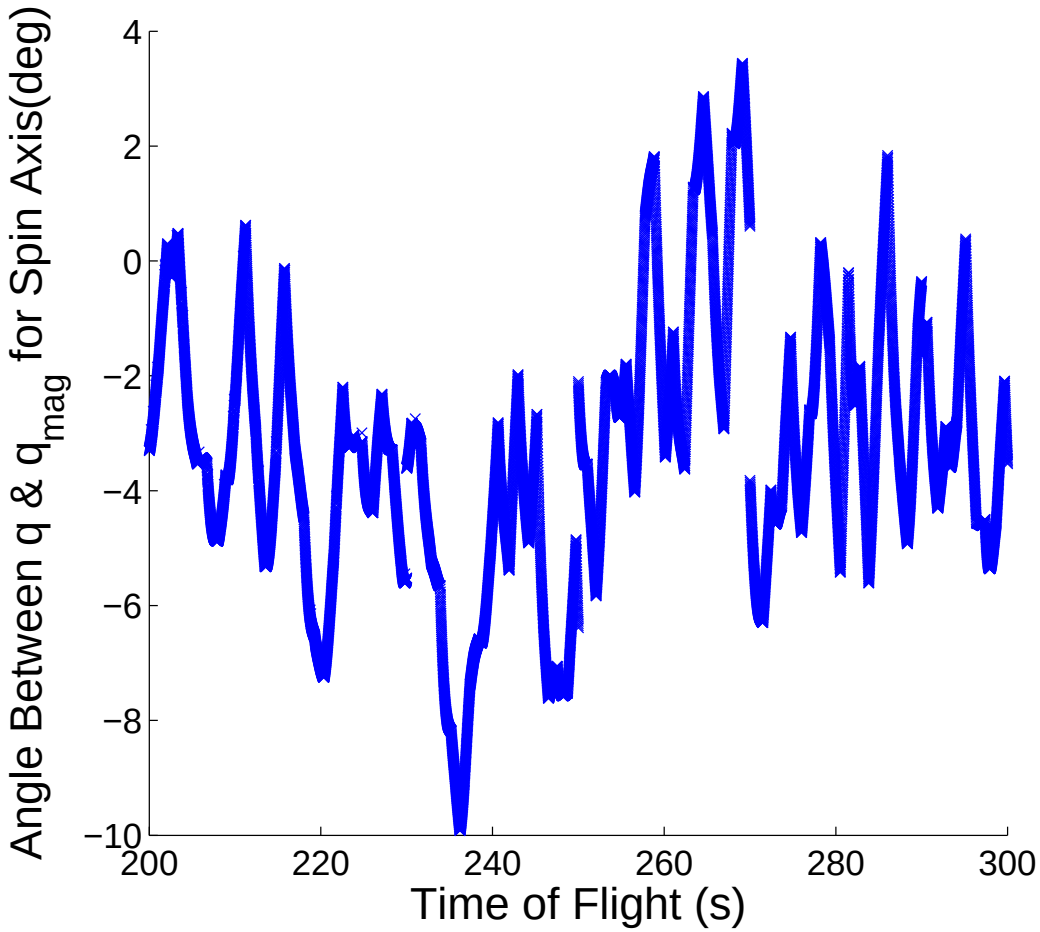

Figure 3.9: GPS vs. magnetometer/HCI angle discrepancy time history about the spin axis.

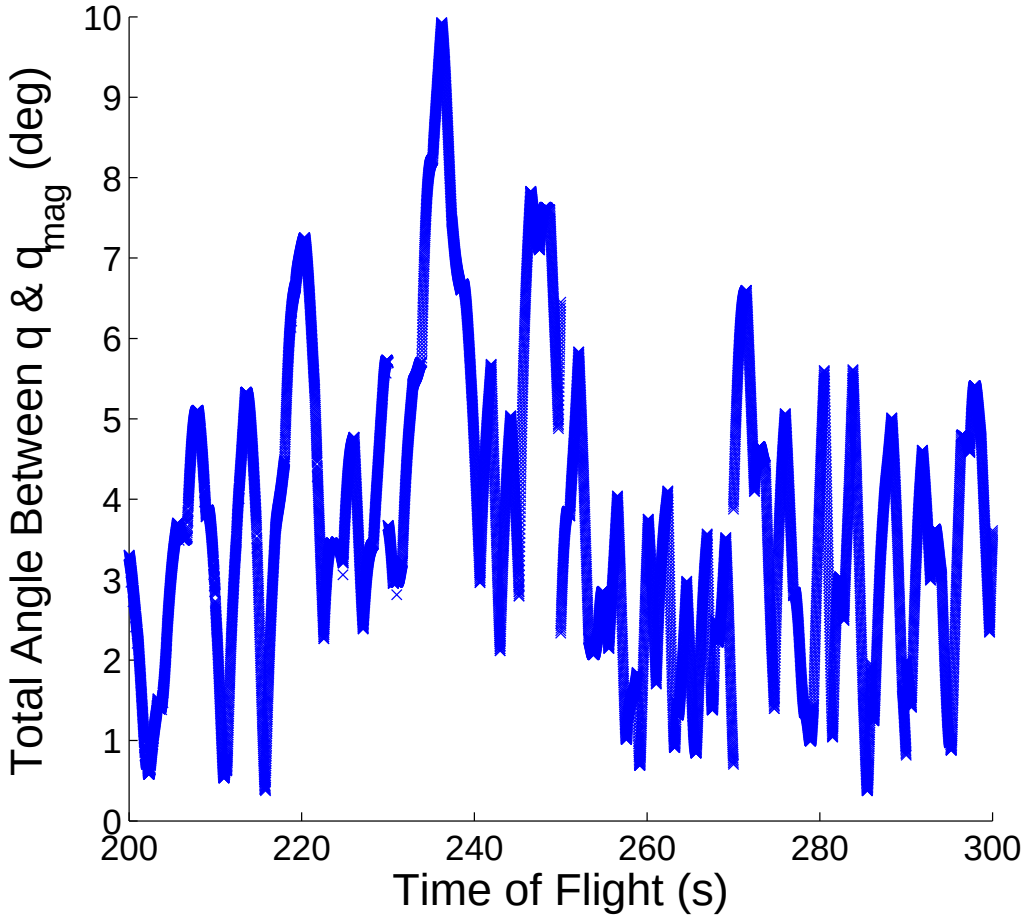

Figure 3.10: Total angular difference time history between GPS and magnetometer/HCI estimates.

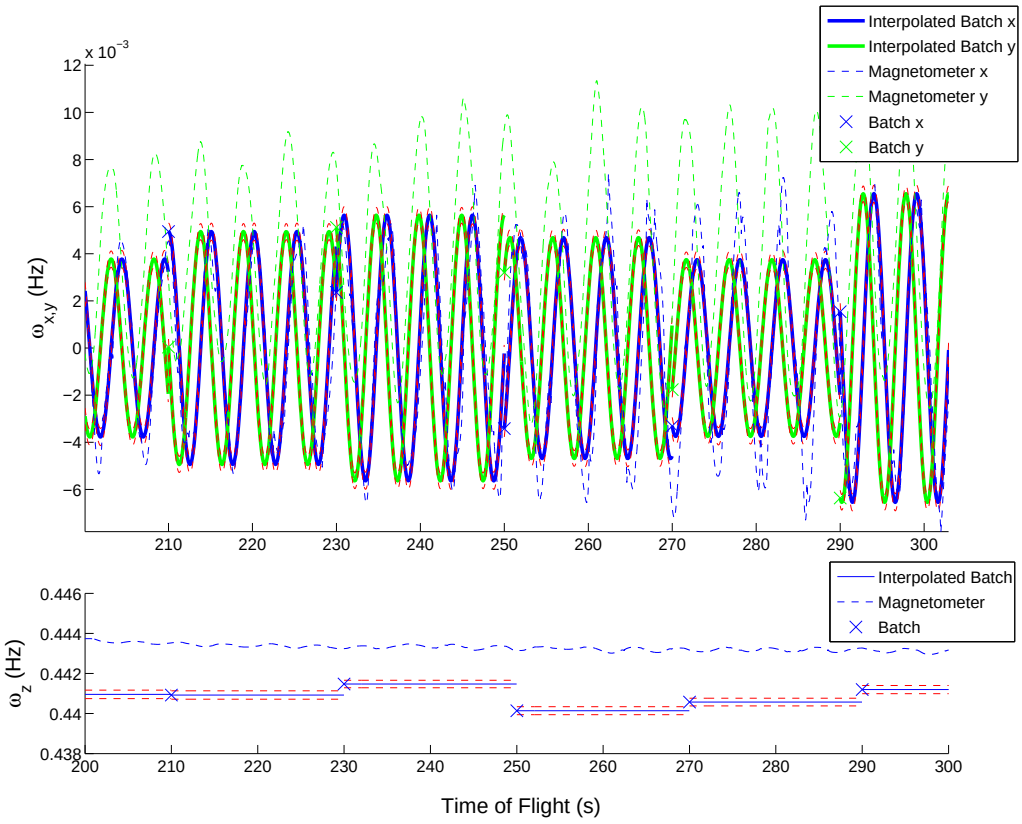

Figure 3.11: Estimated angular velocity component time histories.

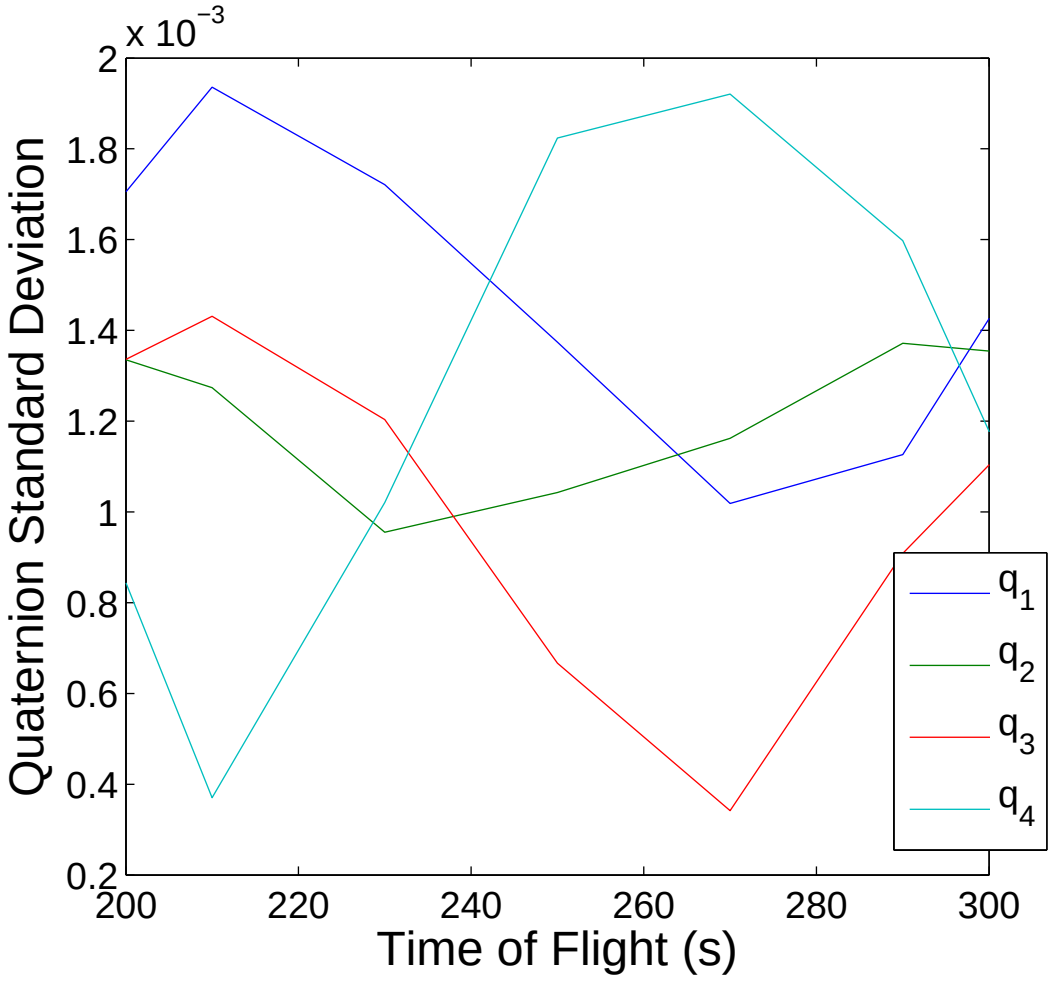

Figure 3.12: Computed quaternion standard deviation time histories from GPS batch filter.

CHAPTER 4

GPS AND IONOSONDE DATA FUSION FOR TOMOGRAPHY OF THE QUIESCENT AND HEATED IONOSPHERE

4.1 Abstract

GPS measurements are combined with ionosonde measurements in an estimation problem for the state of the local ionosphere, both in quiescent and disturbed forms. This estimator has been developed to remotely sense perturbations caused by active heating of the ionosphere above the High Frequency Active Auroral Research Program (HAARP) heater facility in Gakona, AK. The measurement model used by the estimator uses a refractive ray-tracing model of ionosonde signal measurements for the quiescent electron density profile and of dual-frequency GPS signal measurements for both the quiescent and heated profiles. A set of parameter sensitivity calculations augment the ray-tracing solutions in order to facilitate standard estimation-based model inversion calculations. The resulting algorithm determines an optimal parameterization of the ionosphere's electron density profile. Ionosonde data provide information about bottomside ionospheric layers while the HAARP heater is turned off, and trans-ionospheric GPS signal data enable observability of topside structure. The GPS data also allow for a simple characterization of overall profile variations with respect to latitude and longitude, through implicit correlation of slant total electron content (TEC) integrals with the bottom side characterization. The resulting estimate of the full electron density profile is used as a baseline for fitting additional TEC measurements from signals crossing electron density irregularities that are induced when the HAARP heater is active. These data are fit us-

ing an augmented ionosphere model that includes an ellipsoidal field-aligned irregularity of given size, location, and density perturbation profile, but with unknown peak electron density distortion from the quiescent state. Experimental ionosonde and GPS data have been recorded and used to fit the quiescent and disturbed ionosphere models. The results for eight quiescent cases yield an average ionosonde virtual height error of 0.46 km and an average slant TEC error of 3.65 $TECU$ (10^{16} e^{-}/m^{2}), using six GPS satellites. Four disturbed cases demonstrate the viability of the ellipsoidal perturbation model and show close TEC fits, with residual norms under a tenth of a $TECU$.

4.2 Introduction

Improved methods are sought to map the full altitude, latitude, and longitude electron density profile of the ionosphere above a certain locale, especially when the profile has been disturbed by, for example, the High Frequency Active Auroral Research Program (HAARP) heater in Gakona, AK. This capability would benefit areas of atmospheric science where an accurate ionospheric profile is needed. Ionosonde data may be used to generate sufficient representations of the profile below the peak F2 density. Until recently, a measurement-based characterization of the profile above this point could be obtained only from a space-based topside sounder [10] or from an incoherent scatter radar [42]. Dual-frequency GPS data, however, makes topside reconstruction possible, albeit with higher levels of uncertainty.

Previous work has been performed that approximates topside behavior with simple model extrapolation, such as α-Chapman profile matching at the F2 peak

[40][41]. Other authors have fused ionosonde data with GPS TEC measurements, but for the purpose of separating topside and bottomside TEC [43][44] by arithmetic differencing and employing an empirically determined upper transition height. The topside TEC determines the topside profile, but in somewhat of an ad hoc manner, rather than by using optimal estimation techniques directly on both the ionosonde and TEC data.

The present chapter consists of two main contributions, consisting of the development and testing of two techniques. The first is one that optimally fuses GPS TEC data and ionosonde data to estimate a single parameterized local electron density profile. This profile mostly characterizes the vertical profile above the ionosonde/GPS-receiver location, but also includes simple latitude and longitude dependencies. The second method estimates perturbations to the quiescent ionosphere caused by HAARP heating, and is based on slant TEC data for a GPS satellite whose LOS vector pierces the perturbed zone.

The ionosphere refracts radio waves propagating through its magnetoplasma, which affects group delay, carrier phase advance, and the geometric propagation path. The extent of each effect is dependent on wave frequency. Given the design of the dual-frequency GPS signal, the ray-path geometry dependency is negligible, and the other two dependencies can be modeled to first order in terms of the slant total electron content (TEC) along the line-of-sight (LOS) from a ground receiver to a GPS satellite.

One might like to use Ref. [28], however, to show that even the simplest of vertical electron density parameterizations, the Chapman parameterization, is not observable through from GPS slant TEC data alone. By including ionosonde group delay measurements, an optimal approach can be taken to simultane-

ously reconstruct both topside and bottomside. The resulting background density profile can then be used as a baseline. Perturbations from this baseline can be parameterized and the resulting parameters can be estimated as a means of deducing ionospheric irregularities from TEC data during ionospheric disturbances.

The remainder of this chapter is divided into 9 sections. Section 4.3 reviews Hamilton's equations for refractive ray-tracing of radio waves through a magnetoplasma. Section 4.4 describes the parameterization chosen to model the ionosphere and its layers. Sections 4.5 and 4.6 present the algorithms that are used, in conjunction with the Hamiltonian ray-tracing equations, in order to model the ionosonde ray-path observables and the observables' GPS integrals, given the ionospheric electron density parameterization. This section also explains how to compute the partial derivative sensitivities of these observables with respect to electron density profile parameters. Section 4.7 gives the nonlinear optimal estimation methods used to fuse the ionosonde and GPS data in order to estimate the ionospheric parameterization. Section 4.8 presents the ionospheric parameter estimation results that are obtained by this algorithm when using real data from a quiescent ionosphere. Section 4.9 develops a model of the perturbed ionosphere electron density profile that is caused by HAARP heating, and it also develops a method for estimating the perturbation amplitude based on GPS TEC data for a LOS that passes through the perturbed region. Section 4.10 applies this perturbation estimation algorithm to real data from HAARP heating experiments and demonstrates its effectiveness. Section 4.11 presents this chapter's conclusions.

4.3 Hamilton's Equations for Refractive Ray-Tracing in a Magnetoplasma

Refractive ray-tracing of a radio wave with frequency ω entails solving a coupled set of differential equations for the ray path, $\mathbf{r}$, and the wave-front unit direction vector $\boldsymbol{k}$. $k_0 = \omega/c$ is the free-space wave number. The ray-tracing equations are expressed with Hamiltonian formulations that are dependent on $\boldsymbol{r}$, $\boldsymbol{k}$, and k_0 [24]:

$$\frac{d\boldsymbol{r}}{dP'} = \frac{-[dH/d\boldsymbol{k}]^T}{dH/dk_0} \tag{4.1}$$

$$\frac{d\boldsymbol{k}}{dP'} = \frac{[dH/d\boldsymbol{r}]^T}{dH/dk_0} \tag{4.2}$$

where $P' = ct$ is the range-equivalent group delay, chosen as the independent variable for seamless integration through any possible *spitze*'s (singular points of reflection) at the top of an ionosonde path.

The Hamiltonian H is a function that must remain constant along a valid ray path and that can be derived from rewriting the dispersion relation. Two alternate valid forms of the Hamiltonian are used in the present study:

$$H_1(\boldsymbol{r}, \boldsymbol{k}; k_0, \boldsymbol{p}) = \frac{1}{2}\left[\frac{c^2}{\omega^2}\,\|\boldsymbol{k}\|^2 - n^2(\boldsymbol{r}, \boldsymbol{k}; k_0, \boldsymbol{p})\right] \tag{4.3}$$

$$\begin{aligned} H_2(\boldsymbol{r}, \boldsymbol{k}; k_0, \boldsymbol{p}) = real\Bigg\{&\Big[(U-X)U^2 - Y^2U\Big]\frac{k^4}{k_0^4} + X(\boldsymbol{k}\cdot\boldsymbol{Y})^2\frac{k^2}{k_0^4} + \Big[-2U(U-X)^2 \\ &+ Y^2(2U-X)\Big]\frac{k^2}{k_0^2} - X(\boldsymbol{k}\cdot\boldsymbol{Y})^2\frac{1}{k_0^2} + \Big[(U-X)^2 - Y^2\Big](U-X)\Bigg\} \end{aligned} \tag{4.4}$$

Where $X(\boldsymbol{r}; k_0, \boldsymbol{p}) = \omega_{pe}^2/\omega^2$ is the squared ratio of the plasma frequency to the radio wave frequency, and $\boldsymbol{Y}(\boldsymbol{r}; k_0) = Y\hat{\boldsymbol{b}} = (\omega_{ce}/\omega)\hat{\boldsymbol{b}}$ is a vector with magnitude

equal to the ratio of the electron gyrofrequency to the radio signal's frequency and with direction parallel to the local magnetic field. $U = 1 - iZ = 1 - i\nu/\omega$, where ν is the electron collision frequency, is taken to be 1 in the assumed loss-less magnetoplasma. $\boldsymbol{p}$ is a vector of parameters that characterize the electron density profile $N_0(\boldsymbol{r}, \boldsymbol{p})$. It is needed to determine

$$\begin{aligned} \omega_{pe}^2 &= \frac{q_e^2}{m_e \epsilon_0} N_0(\boldsymbol{r}, \boldsymbol{p}) \\ &= 3.183 \times 10^3 N_0(\boldsymbol{r}, \boldsymbol{p}) \end{aligned} \tag{4.5}$$

The specific definition of $\boldsymbol{p}$ will vary depending on the model used. The parameter vector associated with the model of the present algorithm will be discussed in a later section. In Eq. (4.3), n is the phase index of refraction, given by the Appleton-Hartree formula for $Z = 0$ [25]:

$$n^2 = 1 - \frac{X}{1 - \dfrac{Y^2 \sin^2\theta}{2(1-X)} \pm \sqrt{\dfrac{Y^4 \sin^4\theta}{4(1-X)^2} + Y^2 \cos^2\theta}} \tag{4.6}$$

where θ is the angle between wave vector $\boldsymbol{k}$ and magnetic field direction vector $\hat{\boldsymbol{b}}$.

The first Hamiltonian, the one in Eq. (4.3), is appropriate for free space, which may defined as regions where $X \leq X_{min}(Y)$, with the index of refraction significantly different from zero. For ray-tracing through regions with higher electron densities and near-zero index of refraction, that are characterized by $X \geq X_{max}(Y)$, the second Hamiltonian formulation is used, possibly with a sign change. H_2 is especially needed in the neighbourhood of a *spitze*, illustrated in the ray path of Fig. 4.1, because this latter Hamiltonian and the corresponding ray-tracing differential equation in Eqs. (4.1)-(4.2) do not become singular in this case.

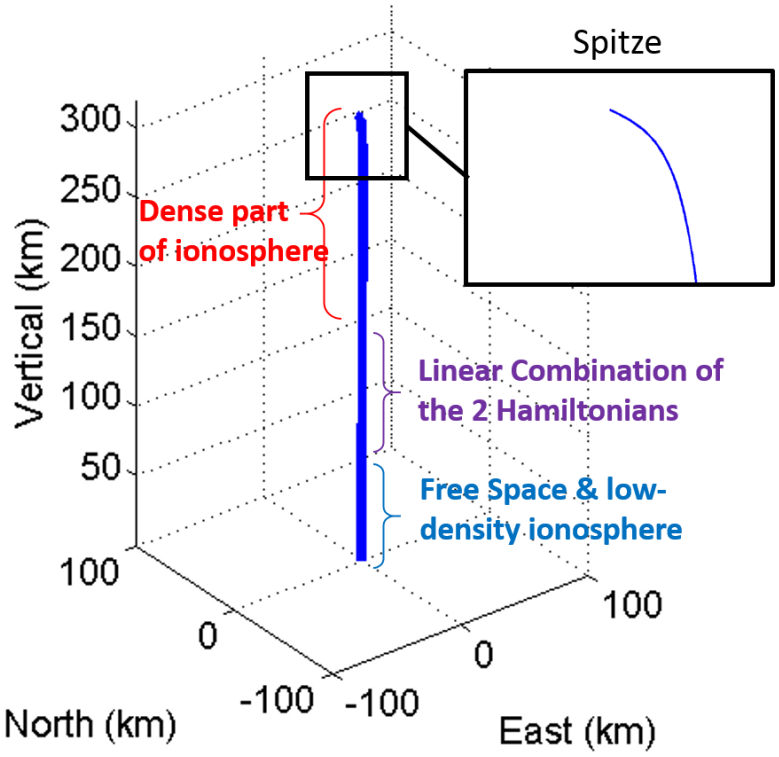

Figure 4.1: Example ray-path solution typical of an ionosonde signal, with a spitze that occurs at the upper reflection point, shown in both normal and magnified views.

The bounds for X differ depending on whether the radio wave is ordinary (O-mode) or extraordinary (X-mode). For some chosen Hamiltonian transition tuning constants a_X and b_X, these limits take the form:

$$X_{min}(Y) = \begin{cases} a_X & \text{if O-Mode or } Y > 1 \\ a_X(1-Y) & \text{Otherwise} \end{cases} \tag{4.7}$$

$$X_{max}(Y) = \begin{cases} b_X & \text{if O-Mode or } Y > 1 \\ b_X(1-Y) & \text{Otherwise} \end{cases} \tag{4.8}$$

where reasonable values for a_X and b_X are 0.1 and 0.25.

In between these two thresholds, i.e., $X_{min}(Y) < X < X_{max}(Y)$, the Hamiltonian used is a linear combination of Eqs. (4.3) and (4.4). Specifically, the transition

Hamiltonian is

$$H_{12} = H_1 \left[1 - \lambda(\alpha)\right] \pm \lambda(\alpha) H_2 \tag{4.9}$$

where the H_2 term is positive if the quadratic polynomial in $(\|\boldsymbol{k}\|/k_0)^2$ that is equivalent to the 2^{nd} Hamiltonian in Eq. (4.4) has a non-negative slope at the root where $(\|\boldsymbol{k}\|/k_0)^2 = n^2$. Otherwise, it is negative. This creates a smooth transition, and avoids the partial derivatives of the two Hamiltonians being canceled out by each other. The new quantities introduced in Eq. (4.9) are defined by

$$\lambda(\alpha) = 10\alpha^3 - 15\alpha^4 + 6\alpha^5 \tag{4.10}$$

$$\alpha = \frac{X - X_{min}(Y)}{X_{max}(Y) - X_{min}(Y)} \tag{4.11}$$

These quantities combine in Eq. (4.9) in a way that causes a smooth transition from the free-space H_1 Hamiltonian to the high-electron-density H_2 Hamiltonian. The transition is a quintic spline in X that transitions from H_1 to H_2 with continuous 0^{th}, 1^{st}, and 2^{nd} derivatives in X.

The ray-tracing equations (4.1)-(4.11) model electro-magnetic wave propagation in the zero-wavelength approximation limit of a lossless magneto-plasma, using the lossless Appleton-Hartree refractive index equation [25]. The Appleton-Hartree formula is used directly in the free-space case of Eq. (4.3), and effectively in the case of Eq. (4.4). Both H_1 and H_2 are non-dimensional, and are zero-valued along a valid ray path, as is their linear combination, H_{12}. H_1, H_2, or H_{12} can be substituted into Eqs. (4.1) and (4.2) for H.

The magneto-ionic parameters X and Y are required to complete this system of equations. Y is dependent on the background magnetic field, which can be obtained from a model such as the International Geomagnetic Reference Field (IGRF). X varies with electron density, the model for which will be covered in

the following section.

4.4 Ionospheric Electron Density Profile Parameterization

There are several potential models that could be used to describe the ionospheric profile. Any candidate $N_0(\boldsymbol{r}, \boldsymbol{p})$ profile must be continuous in $\boldsymbol{r}$ with continuous first and second derivatives, in order to be useable in the ray-tracing equations (4.1)-(4.11). The need for first derivatives is obvious from the form of Eqs. (4.1) and (4.2). The need for second derivatives arises for two reasons. The first is the desire to use a Newton/shooting method to solve the underlying nonlinear two-point boundary value problem for ionosonde ray-tracing. The second reason for using second derivatives arises from the need to solve for Jacobian first-partial-derivative sensitivities of ionosonde ray-tracing solutions with respect to elements of $\boldsymbol{p}$.

One of the simplest and most common ionospheric electron density models that meets the continuity requirement is a superposition of Chapman layers. Other composite electron density models, such as the International Reference Ionosphere (IRI), may also be considered. While these models have various advantages, Booker's "skeleton" model [6] has been chosen for its balance of simplicity, flexibility, and physical meaning. These qualities make it very adaptable to an estimation problem. The Booker model has a spline-like dependency on altitude when considered in the $\ln[N_0(\boldsymbol{r}, \boldsymbol{p})]$ form; it is based on nodes that may be placed at a series of critical altitudes. These nodes, along with transition scale-heights information give rise to a sufficiently smooth vertical profile. An analysis of this model can be found in Ref. [6]. Only one convenient adaptation

will be presented in this chapter.

Booker's vertical electron density profile can be fully described by $3m + 2$ parameters, where m is the number of nodes. By adding two more parameters, this model can be augmented to include a first-order dependence of vertical total electron content (VTEC) on latitude and longitude. The total $3m + 4$ variables comprise the ionospheric Booker profile parameter vector:

$$\boldsymbol{p} = \left[z_1 \quad \ln(z_2 - z_1) \quad \ldots \quad \ln(z_m - z_{m-1}) \quad \ln\left(\frac{1}{Z_{0,1}}\right) \quad \ln\left(\frac{-1}{Z_{m,m+1}}\right) \quad \ln(N_{01}) \right. \tag{4.12}$$
$$\left. \ldots \quad \ln(N_{0m}) \quad \ln(\tilde{z}_1) \quad \ldots \quad \ln(\tilde{z}_m) \quad \frac{\partial \ln(VTEC)}{\partial \phi} \quad \frac{\partial \ln(VTEC)}{\partial \lambda} \right]^T$$

where z_n is the node altitude, N_{0n} is the electron density, and $\tilde{z}_n$ is the transition height for node $n = 1, \ldots, m$. The transition heights control the intervals over which the Booker skeleton slope of the $\ln(N_0)$ versus z function is smoothed to create an analytic function $N_0(z, \boldsymbol{p})$ that yields electron density, given an altitude and a parameterization. $Z_{0,1}$ and $Z_{m,m+1}$ are the first and last Booker skeleton signed scale heights. $\phi = \tilde{\phi} + \Delta\phi$ and $\lambda = \tilde{\lambda} + \Delta\lambda$ denote, respectively, latitude and longitude, which are decomposed into the coordinates of the point of transmission, $(\tilde{\phi}, \tilde{\lambda})$, and the spatial increments $(\Delta\phi, \Delta\lambda)$.

The 3-dimensional point $\boldsymbol{r}(\phi, \lambda, z)$ along the ray path can be expressed in Earth-centred coordinates, and its corresponding latitude and longitude can be calculated using the WGS-84 ellipsoid [14], or another geodetic model. The electron density at $\boldsymbol{r}(\phi, \lambda, z)$, given the local profile parameters, is

$$N_0(z, \boldsymbol{p}) = N_{0,nom} \exp\left[\frac{z - z_{nom}}{Z_{0,1}} + \sum_{n=1}^{m} \left(\frac{\tilde{z}_n}{Z_{n,n+1}} - \frac{\tilde{z}_n}{Z_{n-1,n}} \right) \ln \frac{1 + \exp(z - z_n)/\tilde{z}_n}{1 + \exp(z_{nom} - z_n)/\tilde{z}_n} \right] \tag{4.13}$$

where the nominal altitude is a weighted sum:

$$z_{nom} = \sum_{n=1}^{m} \frac{N_{0n}}{\sum_{j=1}^{m} N_{0j}} z_n \tag{4.14}$$

and the electron density at this altitude is

$$N_{0,nom} = \tilde{N}_{0,nom} \exp\left[\frac{\partial \ln(VTEC)}{d\phi}\Delta\phi + \frac{\partial \ln(VTEC)}{d\lambda}\Delta\lambda\right] \tag{4.15}$$

The remaining unknown parameters of Eq. (4.13) not included in $\boldsymbol{p}$ are the scale heights $Z_{1,2}, ..., Z_{m-1,m}$. These parameters, along with $\tilde{N}_{0,nom}$, can be straightforwardly solved in a well-determined, non-singular, linear system of equations, where the n^{th} equation in this system of m equations in m unknowns is the equivalence, in natural logarithm form, between the Booker-modeled electron density at the n^{th} altitude node and the corresponding value from $\boldsymbol{p}$.

The heavy use of natural logarithms in this version of Booker's model has the effect of ensuring the sign of various quantities or quantity differences that must be positive for a sensible profile. The ionospheric parameter vector $\boldsymbol{p}$ consists of the states that the present algorithm estimates. For each set of parameters, the coupled system of differential equations, Eqs. (4.1) and (4.2), can be solved to give the ray path, the algorithm for which is given in the subsequent section.

4.5 Ionosonde Ray-Tracing Solution for a Given Profile Parameterization

Although the ray-tracing method may be adapted for GPS signals, it does not produce significant discrepancies from a straight-line path due to the high fre-

quency of the GPS L-band signals. The ionosonde ray-tracing case will be presented because the corresponding ray paths can be complicated and require careful solution of Eqs. (4.1)-(4.2).

The ray-tracing differential equations in (4.1) and (4.2) can be assembled into a general state-space model that represents a nonlinear two-point boundary value problem:

$$\frac{d\boldsymbol{x}_{ray}}{dP'} = \boldsymbol{f}(\boldsymbol{x}_{ray}, \boldsymbol{p}) \tag{4.16}$$

$$\text{subject to} \qquad \begin{bmatrix} I & 0 \end{bmatrix} \boldsymbol{x}_{ray}(0) = \begin{bmatrix} \boldsymbol{r}_i \\ 0 \end{bmatrix},$$

$$\begin{bmatrix} 0 & I \end{bmatrix} \boldsymbol{x}_{ray}(P'_f) = \boldsymbol{0},$$

$$\text{and} \qquad H(\boldsymbol{r}_i, \boldsymbol{k}_i; k_0, \boldsymbol{p}) = 0$$

where $\boldsymbol{x}_{ray} = [\boldsymbol{r} \quad \boldsymbol{k}]^T$ is the state vector, $\boldsymbol{r}_i$ is the known initial transmission point of the ray-path, P'_f is the unknown total range-equivalent group delay from the transmitter to the reflection point, and $\boldsymbol{k}_i = [0 \quad I]\boldsymbol{x}_{ray}(0)$ is the initial wave vector. P'_f may be directly compared to virtual height data from the ionosonde. The second constraint in Eq. (4.16) is the terminal boundary condition, which stipulates that, at reflection, the final wave vector is zero.

This nonlinear problem can be solved by first guessing the two unknown components of $\boldsymbol{k}_i$, which dictate its off-zenith and azimuth pointing angles, using the initial condition on the Hamiltonian in Eq. (4.16) in order to compute magnitude of $\boldsymbol{k}_i$, and numerically integrating using Runge-Kutta techniques up to P'_f. A wise guess would be an exactly vertical vector. At the end of the integration, the terminal reflection boundary condition is checked. If it has not been satisfied, then the initial direction of $\boldsymbol{k}_i$ and total group path P'_f are adjusted it-

eratively using Newton's method within this shooting scheme for solving the boundary value problem.

This shooting method requires calculating the sensitivities of the solution to the initial $\boldsymbol{k}_i$ direction variables and to P'_f. The sensitivities to each of the two independent initial directions of $\boldsymbol{k}_i$ can be determined by solving the following initial value problem:

$$\frac{d}{dP'}\left(\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta}\right) = \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}_{ray}}\right|_{\boldsymbol{x}_{ray}(P'),\boldsymbol{p}} \frac{\partial \boldsymbol{x}_{ray}}{\partial \eta} \tag{4.17}$$

where the initial condition $\left.\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta}\right|_0$ must obey

$$\begin{aligned} &\begin{bmatrix} I & 0 \end{bmatrix} \left.\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta}\right|_0 = \boldsymbol{0} \\ &\left.\frac{\partial H}{\partial \boldsymbol{r}}\right|_0 \begin{bmatrix} I & 0 \end{bmatrix} \left.\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta}\right|_0 + \left.\frac{\partial H}{\partial \boldsymbol{k}}\right|_0 \begin{bmatrix} 0 & I \end{bmatrix} \left.\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta}\right|_0 = 0 \end{aligned} \tag{4.18}$$

Given that Eq. (4.18) is 4 equations in 6 unknowns, there are only two possible linearly-independent values of $\left.\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta}\right|_0$: The two that correspond to initial directional variations of $\boldsymbol{k}_i$. Suppose that these are called $\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta_1}$ and $\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta_2}$. The variable η is a placeholder for either of these two directional elements of $\boldsymbol{k}_i$.

The sensitivity differential equations represented by Eqs. (4.17)-(4.18) are integrated twice, once for $\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta_1}$ and once for $\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta_2}$, i.e., once for each independent directional sensitivity of $\boldsymbol{k}_i$. These integrations occur simultaneously with the numerical integration of Eq. (4.16). One can use a $4^{th}/5^{th}$-order Runge-Kutta approximation of the solutions for $\boldsymbol{x}_{ray}$ and $\partial \boldsymbol{x}_{ray}/\partial \eta$.

This ODE integrator has been augmented to take advantage of the fact that the Hamiltonian must remain zero along a valid ray-path. After each $4^{th}/5^{th}$-order Runge-Kutta step, a numerical solver step is taken back towards the $H = 0$

manifold along a direction perpendicular to the local $H = constant$ manifold, in the event that the $H \neq 0$ at the end of the step due to the build-up of numerical error in the integrator.

The Newton's method of solving the shooting problem works with the final two sensitivities $\left.\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta_1}\right|_{P'_f}$ and $\left.\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta_2}\right|_{P'_f}$. It also works with $\left.\frac{\partial \boldsymbol{x}_{ray}}{\partial P'_f}\right|_{P'_f} = f\left[\boldsymbol{x}_{ray}(P'_f), \boldsymbol{p}\right]$. These three sensitivities can be used in the following linearized version of the terminal boundary condition

$$[0 \quad I]\left[\boldsymbol{x}_{ray}(P'_f) + \left.\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta_1}\right|_{P'_f} \Delta\eta_1 + \left.\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta_2}\right|_{P'_f} \Delta\eta_2 + \left.\frac{\partial \boldsymbol{x}_{ray}}{\partial P'_f}\right|_{P'_f} \Delta P'_f\right] = 0. \tag{4.19}$$

Solution of these three linearized terminal boundary conditions for the three unknowns $\Delta\eta_1$, $\Delta\eta_2$, and $\Delta P'_f$ gives the Newton increments for these quantities. The new initial state is $\boldsymbol{x}_{ray}(0) + \left.\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta_1}\right|_0 \Delta\eta_1 + \left.\frac{\partial \boldsymbol{x}_{ray}}{\partial \eta_2}\right|_0 \Delta\eta_2$ with a second-order adjustment to the length of $\boldsymbol{k}_i$ in order to ensure that the initial value of H is zero, and the new terminal group delay is $P'_f + \Delta P'_f$. The iteration is then repeated until the Newton increments $\Delta\eta_1$, $\Delta\eta_2$, and $\Delta P'_f$ all approach zero.

An outer step-size control algorithm has been included in the Newton shooting procedure. It uses the values $\alpha_{step}\Delta\eta_1$, $\alpha_{step}\Delta\eta_2$, and $\alpha_{step}\Delta P'_f$ to update $\boldsymbol{x}_{ray}$ and P'_f for each Newton increment, with the step length α_{step} chosen in the range 0 to 1 in a way that ensures decrease of the sum of the squares of the errors in the terminal boundary condition $[0 \quad I]\,\boldsymbol{x}_{ray}(P'_f) = 0$.

Note that it can be helpful to use various non-dimensionalizations within the numerical integration of Eq. (4.16) and within Newton's method. The version implemented here replaces the ray-path position by its non-dimensionalized position $\boldsymbol{r}/P'_f$, and it replaces the wave vector by its non-dimensionalized form $\boldsymbol{k}/k_0$. The corresponding modifications to the other problem equations are

straightforward.

The chosen $4^{th}/5^{th}$-order Runge-Kutta method has the capability to perform automatic step-size adjustment, but this feature is used only in an outer loop that executes outside the basic Newton/shooting-method solution of the non-linear two-point boundary value problem. Otherwise, the Runge-Kutta step size adjustment could produce spurious results during the Newton step-size adjustment associated with α_{step}. The Runge-Kutta step size adjustment is designed to take small steps where there is a significant numerical integration error indicated by the $4^{th}/5^{th}$-order comparison or by the Hamiltonian deviations from 0 caused by each step prior to re-enforcement of the $H = 0$ constraint.

Other two-point boundary value solvers are also possible substitutes for this portion of the algorithm. The algorithm presented has been found to work very well, usually converging in a few iterations.

As an auxiliary step in ionosonde ray-tracing that is needed in the context of nonlinear estimation, the linearized two-point boundary value problem for the state vector sensitivities with respect to the ionospheric parameters in $\boldsymbol{p}$ needs to be solved. The solution allows for calculating $\partial P'_f/\partial p_j$, which quantifies changes in virtual height with respect to ionospheric parameter p_j, and can ultimately be used to determine the search direction for optimizing each p_j in $\boldsymbol{p}$. For the j^{th} parameter p_j in $\boldsymbol{p}$, the boundary value problem is:

$$\frac{d}{dP'}\left(\frac{\partial \boldsymbol{x}_{ray}}{\partial p_j}\right) = \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}_{ray}}\right|_{\boldsymbol{x}_{ray}(P'),\boldsymbol{p}} \frac{\partial \boldsymbol{x}_{ray}}{\partial p_j} + \left.\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{p}}\right|_{\boldsymbol{x}_{ray}(P'),\boldsymbol{p}} \tag{4.20}$$

where the initial boundary condition at $P' = 0$ is

$$\begin{aligned} &\left[\begin{array}{cc} I & 0 \end{array}\right] \left.\frac{\partial \boldsymbol{x}_{ray}}{\partial p_j}\right|_0 = \mathbf{0} \\ &\left.\frac{\partial H}{\partial \boldsymbol{r}}\right|_0 \left[\begin{array}{cc} I & 0 \end{array}\right] \left.\frac{\partial \boldsymbol{x}_{ray}}{\partial p_j}\right|_0 + \left.\frac{\partial H}{\partial \boldsymbol{k}}\right|_0 \left[\begin{array}{cc} 0 & I \end{array}\right] \left.\frac{\partial \boldsymbol{x}_{ray}}{\partial p_j}\right|_0 = 0 \end{aligned} \tag{4.21}$$

and the terminal boundary condition at P'_f is

$$\left[\begin{array}{cc} I & 0 \end{array}\right] \left\{ \left.\frac{\partial \boldsymbol{x}_{ray}}{\partial p_j}\right|_{P'_f} + \boldsymbol{f}\left[\boldsymbol{x}_{ray}(P'_f), \boldsymbol{p}\right] \frac{\partial P'_f}{\partial p_j} \right\} = 0 \tag{4.22}$$

The linear two-point boundary value problem represented by these equations can be solved using linear shooting techniques much like those that are described in connection with Eqs. (4.17) - (4.19). $\partial P'_f / \partial p_j$ is one of the components of the linear boundary-value solution. It is used directly in the ionosphere nonlinear estimation problem, of which this calculation is a part.

4.6 Calculation of GPS Observables via Integration through a Given Profile Parameterization

Although the same ray-tracing principles may be applied to GPS signals traversing the ionosphere, due to the relatively small deviation of the ray path from the direct (LOS), only perturbations to group delay integrated along the straight-line path between transmitter and receiver need to be considered. These perturbations occur when the group-delay index of refraction differs from 1, and they can be integrated over the path s to give the total effect [19]:

$$\delta P'_{fGPS}(\omega, \boldsymbol{p}) = \int \left[n'(\omega, \boldsymbol{p}) - 1 \right] ds = \int \left[n(\omega, \boldsymbol{p}) + \omega \frac{\partial n(\omega, \boldsymbol{p})}{\partial \omega} - 1 \right] ds \tag{4.23}$$

The total perturbation will depend on the GPS broadcast frequency. The start of the path can be found by determining the satellite's position in its orbit with broadcast ephemerides, and the end of the path is at the receiver. As before, the sensitivities of these variables can also be calculated by numerically integrating them concurrently with Eq. (4.23). $\partial(\delta P'_{fGPS})/\partial \boldsymbol{p}$ is required to optimally estimate the ionospheric electron density distribution parameter vector $\boldsymbol{p}$ using the group delay measurements differenced between the L1 and L2 frequencies.

4.7 Kalman Filter Data Fusion for Optimal Estimation of Ionospheric Parameters of Quiescent Profile

In order to find the optimal estimates of the profile parameters in $\boldsymbol{p}$, data from the ionosonde and dual-frequency GPS are fused using a Square Root Extended Kalman Filter (SREKF). The present implementation of the Booker parameterization often leads to an ill-conditioned estimation problem. The SREKF is known to be a numerically stable form of the extended Kalman filter (EKF), which is itself an approximation of the optimal Bayesian filter for nonlinear estimation problems. It has been used successfully for optimizing the quiescent profile where other algorithms have failed. For details of the EKF and Square Root Information Filter, see Refs. [2], [5], and [39]. Only the models and initialization procedures for the filter will be given.

4.7.1 Measurement and Dynamic Models

The ray-traced ionosonde virtual heights and the integrated GPS L1/L2 frequency-differenced group delays, which are all functions of the state vector $\boldsymbol{p}$, are stacked as the vector of measurements that define the nonlinear measurement function $\boldsymbol{h}$:

$$\boldsymbol{h}(\boldsymbol{p}, DCB_{RX}) = \begin{bmatrix} P'^{1}_{f}(\omega_1, \boldsymbol{p}) \\ \vdots \\ P'^{n_I}_{f}(\omega_{n_I}, \boldsymbol{p}) \\ \Delta P'^{1}_{fGPS}(\boldsymbol{p}, DCB_{RX}) \\ \vdots \\ \Delta P'^{n_{GPS}}_{fGPS}(\boldsymbol{p}, DCB_{RX}) \\ N_0(0, \boldsymbol{p}) \\ \lim_{z\to\infty} N_0(z, \boldsymbol{p}) \end{bmatrix} \tag{4.24}$$

where DCB_{RX} is the unknown inter-frequency GPS receiver bias. The quantities $P'^{i}_{f}(\omega_i, \boldsymbol{p})\ ;\ i = 1, \ldots, n_I$ are the virtual heights from the ionosonde, modeled by numerical integration of Eq. (4.16) for each sounding frequency ω_i. The frequency-differenced GPS group delay observables are:

$$\begin{aligned} \Delta P'^{j}_{fGPS}(\boldsymbol{p}, DCB_{RX}) &= P'^{j}_{fL2} - P'^{j}_{fL1} \qquad ; \quad j = 1, \ldots, n_{GPS} \\ &= \Delta\Psi^{j}(\boldsymbol{p}) - c(DCB^{j} + DCB_{RX}) + \Delta w^{j}_{\Psi} \end{aligned} \tag{4.25}$$

where $\Delta\Psi^{j}(\boldsymbol{p}) = \delta P'^{j}_{fGPS}(\omega_{L2}, \boldsymbol{p}) - \delta P'^{j}_{fGPS}(\omega_{L1}, \boldsymbol{p})$ is the difference in group delay between the L1 and L2 GPS frequencies for satellite j as modeled using Eq. (4.23). These differential quantities are used for the GPS data because differencing removes the unknown effects of satellite to receiver geometry, receiver clock error, and satellite clock error. These differences only contain total electron content (TEC) information. Moreover, it is assumed that the differential

pseudoranges have undergone carrier-smoothing, via one of the many existing methods [18][21], such that the measurements entering into Eq. (4.25) are both precise and ambiguity-free.

The last two elements of the measurement vector in Eqn. 4.24 serve as soft constraints on the shape of the electron density profile, ensuring that the density on the Earth's surface and that at a very high altitude are both zero. These soft constraints help to mitigate some divergence issues during optimization.

Equation (4.25) includes real-world effects on its right-hand side that model the ways in which actual receiver-generated frequency-differenced group delay varies from its theoretically modeled value in $\Delta\Psi^j(\boldsymbol{p})$. These differences arise from various instrument biases and noise terms. DCB^j is satellite j's inter-frequency differential code bias, which can be obtained from the Center for Orbit Determination in Europe (CODE) [30]. DCB_{RX}, as before, is the differential code bias of the receiver. Δw_Ψ^j is a general white measurement noise term for the frequency-differenced pseudorange of the j^{th} satellite.

DCB_{RX} is unknown and needs to be estimated along with $\boldsymbol{p}$. Therefore, an augmented state vector $\boldsymbol{x}_k = [\boldsymbol{p} \quad DCB_{RX}]_k^T$ is constructed for the SREKF measurement model, which takes the final form:

$$\boldsymbol{y}_k = \boldsymbol{h}(\boldsymbol{x}_k) + \boldsymbol{w}_k \tag{4.26}$$

where $\boldsymbol{y}_k$ is a stacked vector at filter step k consisting of the raw ionosonde virtual height data and frequency-differenced, carrier-smoothed pseudorange data from GPS. The zero-mean Gaussian white-noise vector $\boldsymbol{w}_k$ is comprised of the ionosonde virtual height measurement noise components for the sounding frequencies $\omega_1, ..., \omega_{n_I}$ and the frequency-differenced GPS pseudorange measurement noise components $\Delta w_\Psi^1, ..., \Delta w_\Psi^{n_G PS}$ at time k. This noise vector has an asso-

ciated covariance matrix

$$R = \text{diag}\left(\sigma_1^2, \ldots, \sigma_{n_I}^2, \sigma_{\Psi 1}^2, \ldots, \sigma_{\Psi n_{GPS}}^2, \sigma_{C0}^2, \sigma_{C\infty}^2\right) \tag{4.27}$$

which has diagonal entries comprised of the variances of the measurement noise terms. Ionosonde precision is typically within a few kilometres, and therefore a standard deviation around $\sigma = 1\ km$ would be a good choice. Similarly, each $\sigma_\Psi = 0.01\ m$ are reasonable values to use for the measurement uncertainty in frequency-differenced, carrier-smoothed GPS pseudorange. σ_{C0} and $\sigma_{C\infty}$ are the standard deviations of the soft constraints at $z = 0$ and $z \to \infty$, respectively, and act essentially as tuning parameters that have been set to 1.

This effort seeks to estimate snapshots of the ionospheric background electron density profile that span the order of a few minutes. Therefore, the dynamic model for the parameter vector $\boldsymbol{p}$ is assumed to be a random walk model:

$$\boldsymbol{p}_{k+1} = \boldsymbol{p}_k + \boldsymbol{v}_k \quad ;\ v \sim \mathcal{N}(0, \epsilon I) \tag{4.28}$$

where ϵ is on the order of 10^{-6}. This small value of white process noise covariance has the effect of modeling the parameters in $\boldsymbol{p}$ as being nearly constant over the data spans of interest. If a longer span is needed, a Markov-process dynamic model for the $\boldsymbol{p}$ vector like that in Ref. [28] might prove useful. Note that this value of ϵ still holds with logarithmic states. For example, if $p_i = \log(\tilde{p}_i)$ is one such state, then its noise variance $E[(p_{i,k+1} - p_{i,k})^2] = E[\log^2(\tilde{p}_{i,k+1}/\tilde{p}_{i,k})]$ needs to be near zero in order for $\tilde{p}_{i,k+1}/\tilde{p}_{i,k} \approx 1$.

The inter-frequency GPS receiver bias is modeled as being constant. Thus, its dynamic model takes the form:

$$DCB_{RX,k+1} = DCB_{RX,k} \tag{4.29}$$

4.7.2 Filter Implementation

Eqn. (4.26) constitutes the measurement equation, and Eqns. (4.28) - (4.29) are the dynamic model equations in a standard SREKF [2] [5] [39]. While developing the algorithms of this Chapter, several variations of the Levenberg-Marquardt batch filter were also tested, including ones deemed robust to initial guesses [47]. The SREKF was chosen, as this framework allows for direct weighting of the filter states through the dynamic models uncertainties. This helps keep the ionospheric shape feasible despite unobservabilities, as discussed in Section 4.8. A typical sampling period for this implementation is $1s$, and the filter typically processes sets of 20 samples. The filter converges within 5 to 10 sample sets from the initial guess, indicated by negligible increments of the states, using the standard measurement model linearizations of an SREKF. However, the filter still is very sensitive to the initial guess as well as the initial state covariances. Upon starting out with an initial guess of the states that is moderately far from a good estimate, or with state covariances that are too high, the filter produces unfeasible profiles that are similar to the examples of poor results (Figs. 4.4 - 4.6) in Section 4.8. Viable initial guesses for the states can generally be achieved by implementing the procedure in the following section, Section 4.7.3, using node heights $z_1, ..., z_n$ that correspond to inflection points in the ionosonde-generated profile, i.e., at the F2-peak height, F-1 peak height, E-peak height, E-valley, the approximate upper transition height, and the height approximately one E-layer half-width below the E-peak. Moving these nodes or using more nodes between these critical points leads to the aforementioned poor profile results, which in turn causes singular errors during the numerical integration process. Initial covariances were chosen via experimental tuning, and conservative values were determined to be best for convergence, also as

discussed in Section 4.8. An iterated SREKF may be an even more suitable option to deal with the heavy nonlinearities of this problem in the future, however, it would be best if parallelization or some other compensation for the computational overhead is implemented. The next subsection presents the method with which the SREKF is initialized.

4.7.3 Filter Initialization

The initial guess for the ionospheric parameter vector $\boldsymbol{p}$ is generated by utilizing the auto-scaled numerical real height function from the ionosonde's software, ARTIST-5. Figure 4.2 is an example ionogram processed by ARTIST-5. The real heights are used in a Levenberg-Marquardt algorithm that minimizes the cost

$$J(\boldsymbol{p}) = \frac{1}{2}\sum_{i=1}^{M}\left[N_{Ri} - N_0(z_i, \boldsymbol{p})\right]^2 + \frac{1}{2}w\sum_{n=1}^{m}\left[z_{ng} - z_n(\boldsymbol{p})\right]^2 \tag{4.30}$$

where N_{Ri} for $i = 1, \ldots, M$ is the electron density at the real height z_i determined by the ionosonde's software. This is compared to the aforementioned Booker density model $N_0(z_i, \boldsymbol{p})$. z_{ng} is an input guess of the n^{th} node altitude z_n. Examples of node placement can be found in Ref. [6] and in Fig. 4.3, where the blue curve denotes the logarithmic skeleton profile that simply joins the nodes without smoothing, i.e., only uses the z_n and N_{0n} terms of $\boldsymbol{p}$. To avoid a severely ill-conditioned problem, the nodes are typically the approximate inflection points in the real density function. The red dashed curve is the full logarithmic profile that incorporates the $Z_{0,1}$, $Z_{m,m+1}$, and $\tilde{z}_n$ terms as well. w is a positive scalar that weights the cost contributions of the differences between each final optimized altitude node value and z_{ng}. First guesses of other ionospheric parameters in the

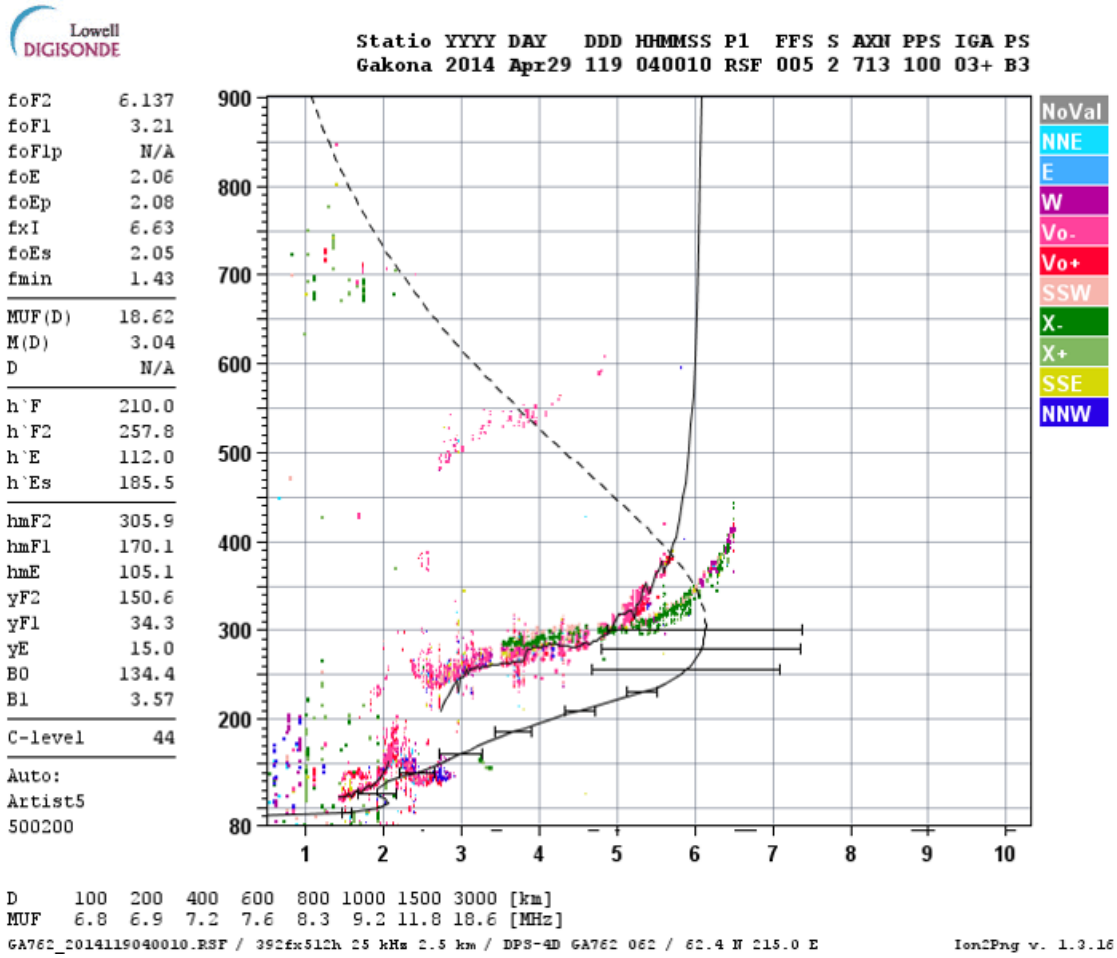

Figure 4.2: Example ionogram with virtual and real height profiles auto-scaled by ARTIST-5. Virtual heights are automatically fitted, often with some error, to the lowest echoes.

$\boldsymbol{p}$ vector used to initialize this Levenberg-Marquardt algorithm are produced by polynomial fitting and interpolation.

The resulting guess $\boldsymbol{x}_g = [\boldsymbol{p} \quad \boldsymbol{0}]^T$, optimized to minimize the cost function in Eqn. (4.30) for $\boldsymbol{p}$, is typically close to the optimal estimate of $\boldsymbol{x}$ produced by the SREKF of the previous section for the bottom-side ionosphere, i.e. below the peak density altitude. However, the top-side carries higher uncertainty because only ionosonde information is incorporated into $\boldsymbol{x}_g$, and the topside distribution from the ionosonde data is based on a number of assumptions that are only

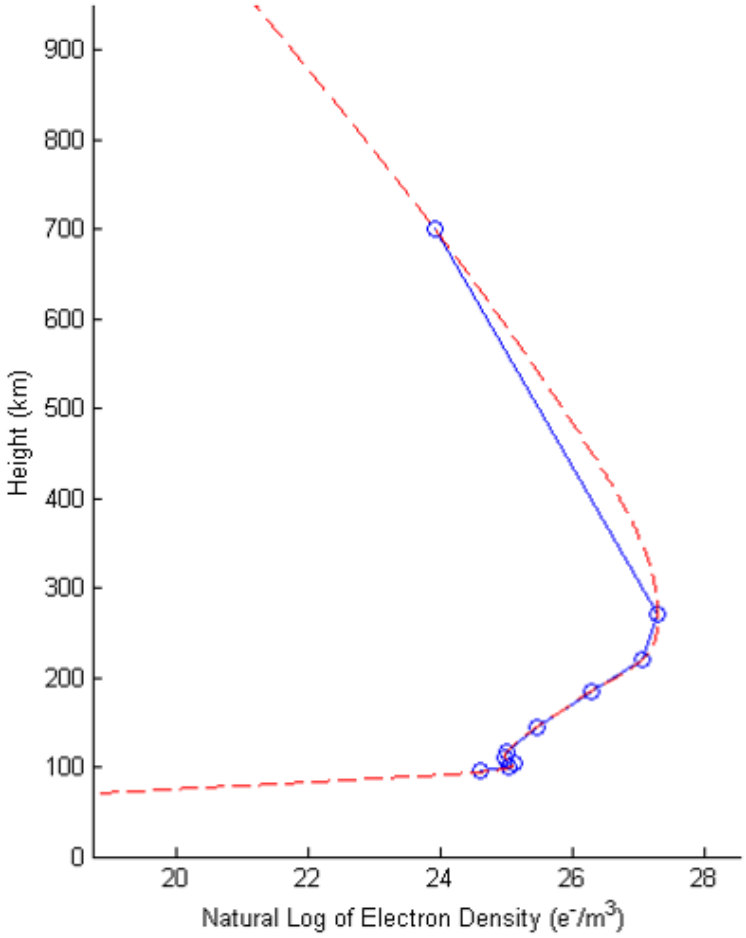

Figure 4.3: Example log of density vs. height function with nodes.

rough approximations of reality. This discrepancy can be accounted for by a higher variance attributed to the initial node height and electron density for the node in the top-side and attributed to the N_{Ri} vs. z_i ionosonde values for i values from the topside model. The initial state covariance matrix would then be

$$P_{xx} = \text{diag}\left(\sigma^2_{z_1}, \sigma^2_{z_2-z_1}, \ldots, \sigma^2_{z_m-z_{m-1}}, \sigma^2_{Z_{0,1}}, \sigma^2_{Z_{m,m+1}}, \sigma^2_{N_{01}}, \right. \\ \left. \ldots, \sigma^2_{N_{0m}}, \sigma^2_{\tilde{z}_1}, \ldots, \sigma^2_{\tilde{z}_m}, \sigma^2_{dVTEC_\phi}, \sigma^2_{dVTEC_\lambda}, \sigma^2_{DCB}\right) \tag{4.31}$$

where each σ is the associated standard deviation for each state. Then σ_{z_m} and

$\sigma_{N_{0m}}$ in particular may be made a few orders of magnitude larger than the other $\boldsymbol{p}$ states.

4.8 Ionospheric Parameterization Kalman Filter Results

This section presents results that have been produced by applying the ionosonde-GPS SREKF algorithm on ionosonde virtual height data and dual-frequency GPS pseudorange data for quiescent conditions. The data were collected on April 29, May 29, May 30, and June 9, 2014 UTC, at the High Frequency Active Auroral Research Program (HAARP) at Gakona, Alaska, during an experiment campaign. The quiescent data immediately before and after experimental heating times have been analyzed in order to resolve a baseline for perturbations. The virtual heights have been extracted from the HAARP digisonde, developed by the University of Massachusetts at Lowell, and the multiple-echo ionograms are fitted to the lowest echo with a combination of ARTIST-5 software and manual tweaking with SAO Explorer software, the latter to compensate for any ARTIST-5 errors. GPS data are retrieved from Miami University's Novatel OEM4 receiver at the HAARP facility. The Novatel receiver is capable of tracking P(Y) code, and therefore it became unnecessary during the course of the campaign to concentrate only on L2C-capable GPS satellites at the HAARP receiver location.

Each quiescent case is divided into 20 sets of samples at 1 s intervals, where each measurement set at time $k = 1, ..., 20$ contains the frequency-differenced pseudoranges of several GPS satellites, as well as a set of ionosonde virtual heights recorded at the start of the first batch. The ionosonde sounding fre-

quencies range from about $1MHz$ to each peak $f0F2$ frequency, and the corresponding virtual heights are fitted to the refractive ray-tracing model outputs $P'_f(\omega_i, \boldsymbol{p}_k)$ at each k.

The ionosonde and GPS signals traverse a quiescent ionosphere. They are used to estimate a parameterization of the background profile prior to estimation of irregularities in the electron density profile that will be caused by HAARP heating. The typical initial state standard deviations used are given in Table 4.1, for $m = 6$ Booker-profile nodes. The seemingly small values have been chosen in part to reduce wild behaviour during profile optimization, which can yield faults in integration, most likely due to the highly nonlinear and the logarithmic nature of the states. As previously mentioned, the top node can be assigned larger uncertainties, but one must still be conservative with the standard deviation, as evidenced in Figs. 4.4, 4.5, and 4.6, where "high" values, implying σ's on the order of 1 $ln(m)$, 1 $ln(e^-/m^3)$, or 1 $ln(m^{-1})$ lead to severely improbable profile shapes. In Fig. 4.4, for example, the top node has moved too far, causing the lower half of the F-region to jut out excessively past $f0F2$. It has also caused the profile to approach an infinite plasma frequency at the top, where it should be approaching zero. Figure 4.5 contains a profile that has been completely inverted due to large movement among all the nodes. Another contributing factor to erratic profile outputs may be that the valley between the E- and F-regions is unobservable with respect to both the ionosonde and to GPS TEC measurements. The unobservability seems most evident in Fig. 4.6, where it is manifested in the incorrect E-layer critical frequency $f0E$ and an overall flattened E-region, the corresponding nodes of which have been given too much freedom.

State Standard Deviation	Value
σ_{z_1}	$1m$
$\sigma_{z_2}, \ldots, \sigma_{z_5}$	$10^{-6}ln(m)$
σ_{z_6}	$0.1ln(m)$
$\sigma_{Z_{0,1}}, \sigma_{Z_{6,7}}$	$10^{-2}ln(m^{-1})$
$\sigma_{N_{01}}, \sigma_{N_{03}}$	$10^{-6}ln(e^-/m^3)$
$\sigma_{N_{02}}$	$10^{-2}ln(e^-/m^3)$
$\sigma_{N_{04}}$	$10^{-3}ln(e^-/m^3)$
$\sigma_{N_{05}}$	$0.1ln(e^-/m^3)$
$\sigma_{N_{06}}$	$1ln(e^-/m^3)$
$\sigma_{\tilde{z}_1}, \ldots, \sigma_{\tilde{z}_6}$	$10^{-2}ln(m^{-1})$
$\sigma_{dVTEC_\phi}, \sigma_{dVTEC_\lambda}$	$1rad^{-1}$
σ_{DCB}	$10^{-9}s$

Table 4.1: Initial SREKF standard deviations for each state.

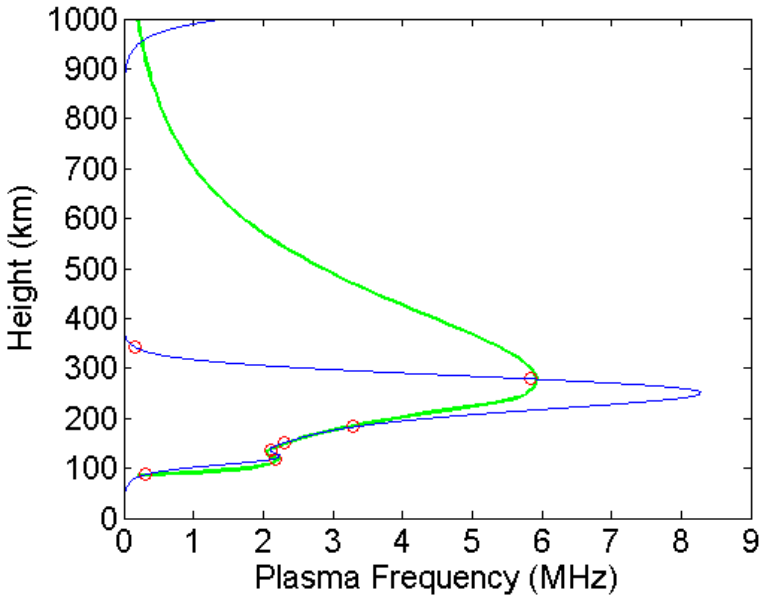

Figure 4.4: Wild profile behaviour when σ_{z6} has been set too high. Red circles denote the nodes, the blue line is the estimated profile, and the profile generated by ARTIST-5 is in green.

Figures 4.7 and 4.8 show a typical fit of the data that were collected on April 29, 2014, produced by the SREKF and its final parameterization $\boldsymbol{p}$ at the end of the 20 s filtering run, at the end of which the increments to $\boldsymbol{p}$ are small. This run used data from 5 GPS satellites and 25 ionosonde sounding frequen-

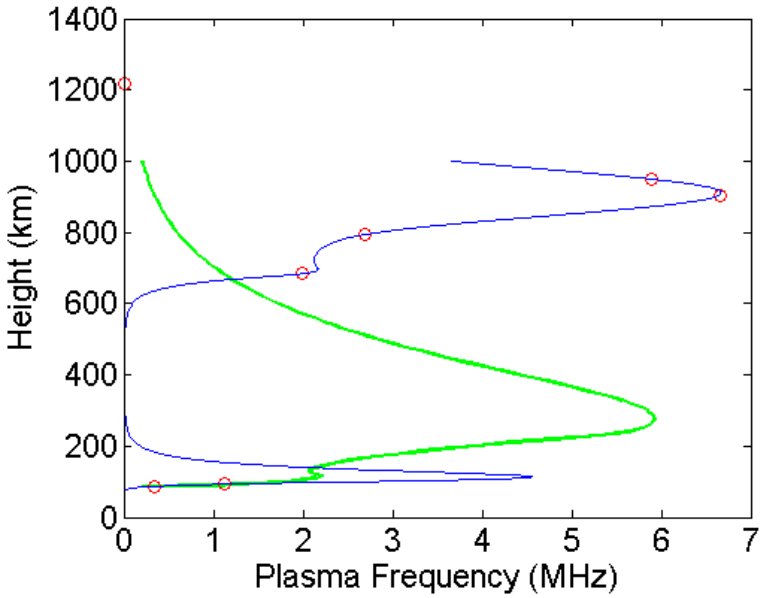

Figure 4.5: Wild profile behaviour when σ values for all states have been set too high. Red circles denote the nodes, the blue line is the estimated profile, and the profile generated by ARTIST-5 is in green.

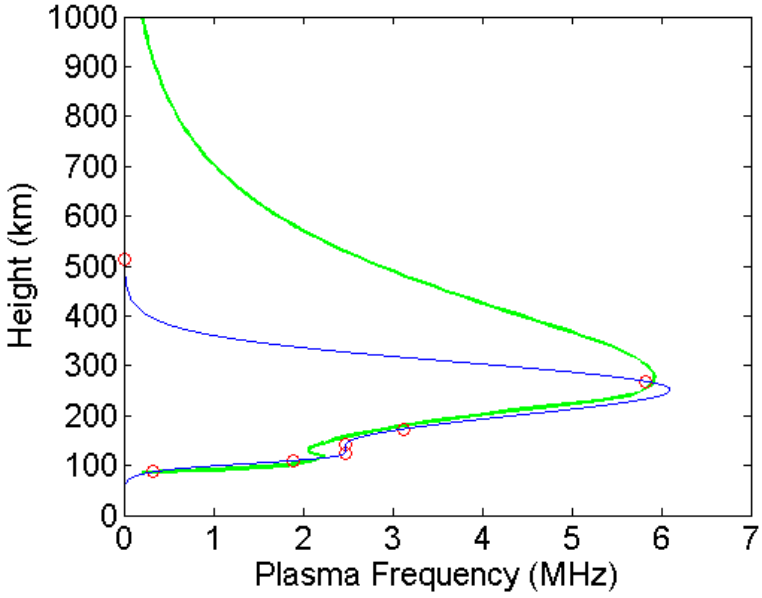

Figure 4.6: Undesirable profile behaviour when σ values for electron density states about E-region have been set too high. Red circles denote the nodes, the blue line is the estimated profile, and the profile generated by ARTIST-5 is in green.

cies ranging from 1.45 to 5.8 MHz. The discrepancies between the model and ionosonde data virtual heights are within a few kilometres for most points on Fig. 4.7. Note that it is usually difficult to fit the more abruptly curving portions of the ionosonde data without adding more nodes; however, an increase in the number of nodes causes the estimation problem to become more unobservable. Data-model differences, given in $TECU$ (10^{16} e^{-}/m^{2}), for the GPS slant TEC values are significant, as shown on Fig. 4.8. The ionospheric profile model for differential group delay tends to be ill-fitted for satellites at lower elevations. For the eight quiescent cases, more weight has been attributed to the satellites that are at higher elevations, as the TEC data from high-elevation satellites are likely to be more accurate. The final estimated profile is presented in Fig. 4.9, along with the ARTIST-5 automatic profile from the ionosonde. The last two ionospheric parameters, the sensitivities of the logarithms of VTEC with respect to latitude and longitude, have values $\partial \ln(VTEC)/\partial\phi = -2.20\ rad^{-1}$ and $\partial \ln(VTEC)/\partial\lambda = -0.76\ rad^{-1}$. The corresponding VTEC percentage change per 5 degrees of latitude and longitude are, respectively, 20% and 18%, which are somewhat high, and may not really be representative of the true ionosphere. The most obvious discrepancy between the estimated profile and the ARTIST-5 output is in the top-side profile, due to the additional information from the GPS TEC measurements. The numerical RMS errors for this case, as well as the rest of the eight cases, are laid out in Table 4.2.

Figures 4.10, 4.11, and 4.12 are plots from another example case, from May 29, 2014, which incorporates more low-elevation satellites, and shows poorer TEC fits. The sensitivities of the logarithm of VTEC to latitude and longitude for the estimated profile are $\partial \ln(VTEC)/\partial\phi = 0.10\ rad^{-1}$ and $\partial \ln(VTEC)/\partial\lambda = 0.13\ rad^{-1}$.

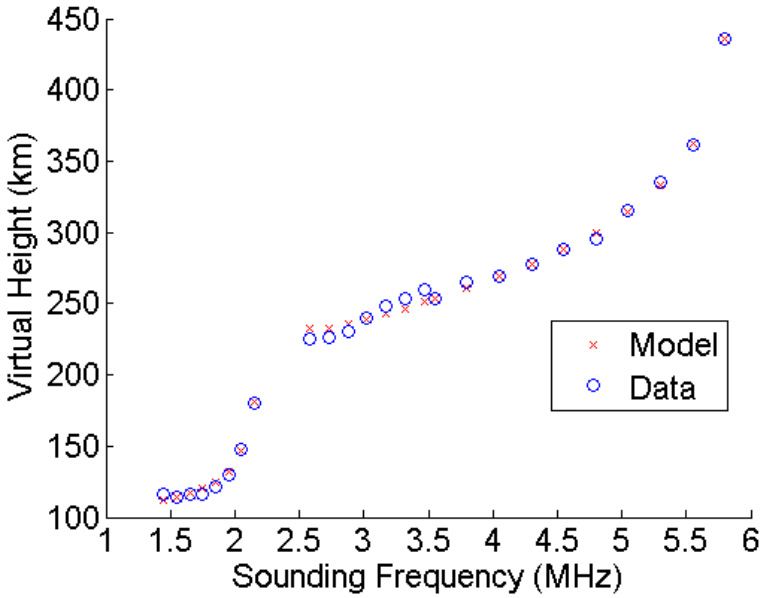

Figure 4.7: Final Kalman filter fitting of ionosonde virtual heights from April 29, 2014 case.

Case Date (UTC)	Case Time (UTC)	RMS Virtual Height Error (km)	RMS TEC Error (TECU)
4/29/2014*	4:07:40	3.79	1.74
4/29/2014*	4:52:40	5.82	6.73
5/29/2014	4:45:10	4.85	8.37
5/29/2014	5:15:10	8.09	6.08
5/30/2014	4:37:40	25.18	8.45
5/30/2014	5:07:40	18.61	5.97
6/9/2014	4:00:10	12.75	6.24
6/9/2014	4:30:10	3.37	5.06

Table 4.2: Table of virtual height and root mean square (RMS) TEC errors of six satellites for eight Booker profile fits. (*Except for 4/29/2014, which uses five satellites).

4.9 Ionospheric Irregularity Parameterization with GPS Signals

Several heating experiments have been conducted using the HAARP Ionosphere Research Instrument (IRI), a high-frequency (HF) pump. This antenna

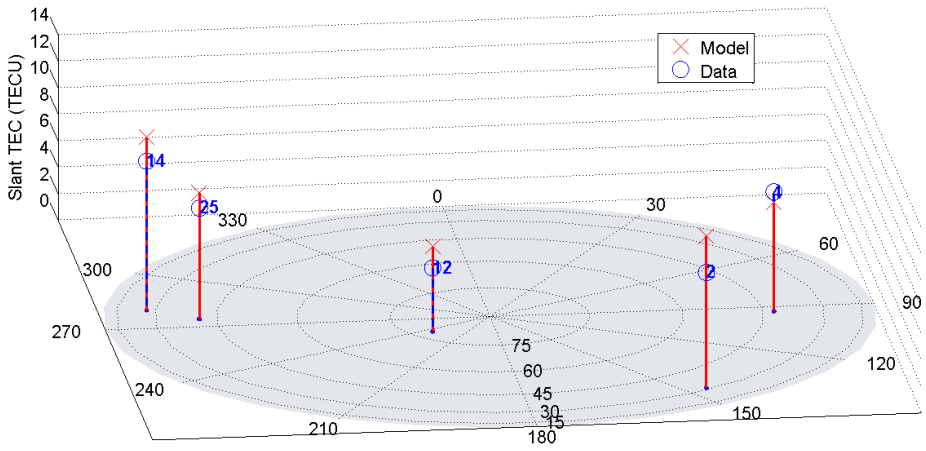

Figure 4.8: Final Kalman filter fitting of slant TEC for 5 GPS satellites for April 29. The satellite locations are shown on an elevation-azimuth polar sky-plot, with PRN numbers in blue.

system transmits up to a $3.6MW$ signal into the ionosphere to excite the electrons in the magneto-plasma, which can cause electron density depletion or enhancement irregularities. Whether the perturbation is enhancement or depletion depends on a myriad of environmental factors, for which there is currently no predictive model.

The irregularities are known to be aligned with the local Earth magnetic field, and they are termed field-aligned irregularities (FAI). Their field-alignment is caused by electron diffusion [15] along the Earth's magnetic field lines. These irregularities have scale sizes above the Fresnel length, and may be probed by dual-frequency GPS signals to ascertain their sizes and shapes as refractive model parameters. Typical FAI's span a few kilometres in width, and may stretch hundreds of kilometres in length along the Earth's magnetic field. The perturbed zone's width is governed by the width of the HAARP heating zone, and its length by diffusion and ion/electron recombination physics. The experimental strategy for creating these disturbance bubbles in the quiescent

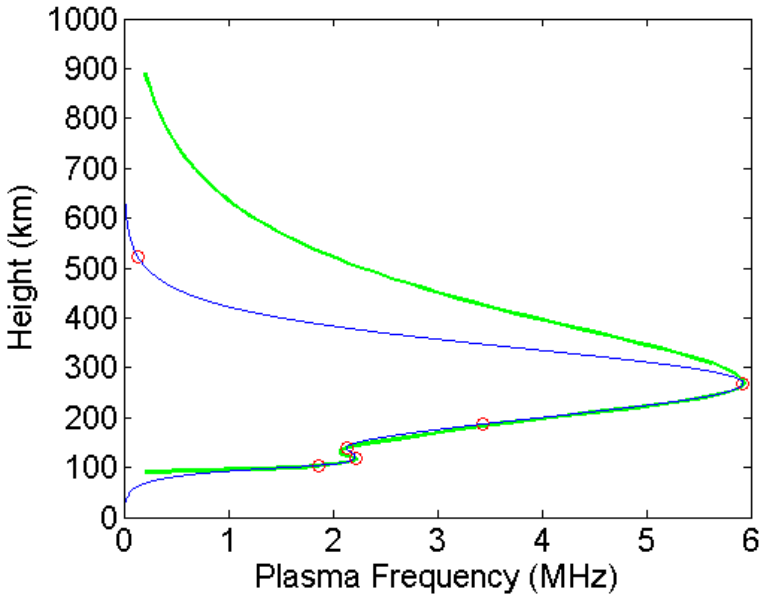

Figure 4.9: Final Booker profile (blue) using six nodes (red circles), and ARTIST-5's profile (green), for April 29, the former as estimated by the SREKF and the latter synthesized as an output of the ionosonde, all for a quiescent ionosphere.

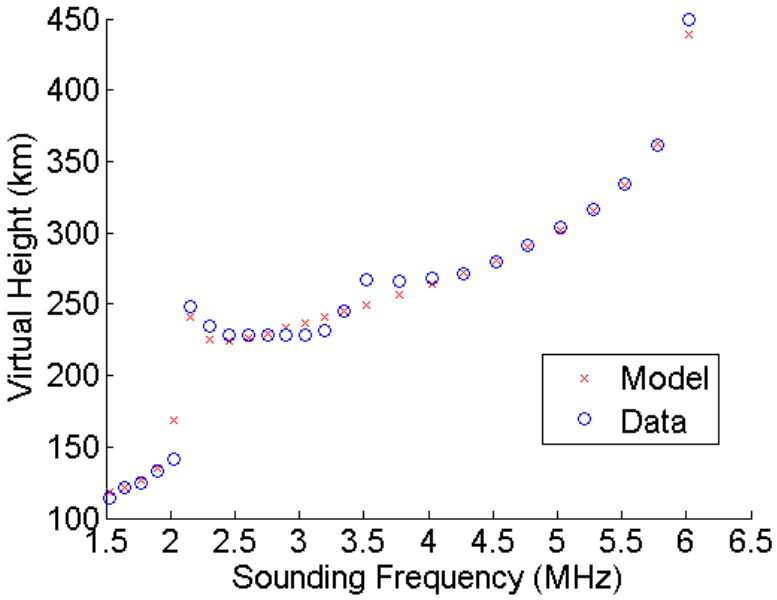

Figure 4.10: Final Kalman filter fitting of ionosonde virtual heights, for data collected on May 29, 2014.

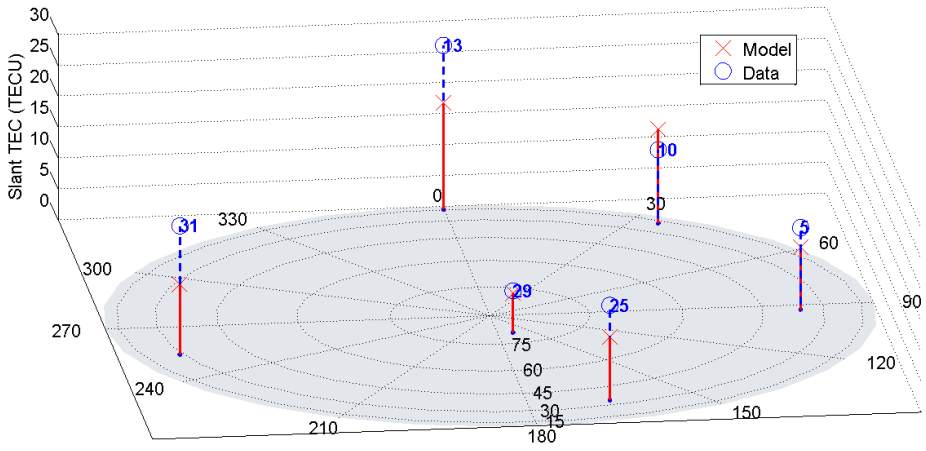

Figure 4.11: Final Kalman filter fitting of slant TEC for 6 GPS satellites. The satellite locations are shown on an elevation-azimuth polar sky-plot, with PRN numbers in blue, for May 29.

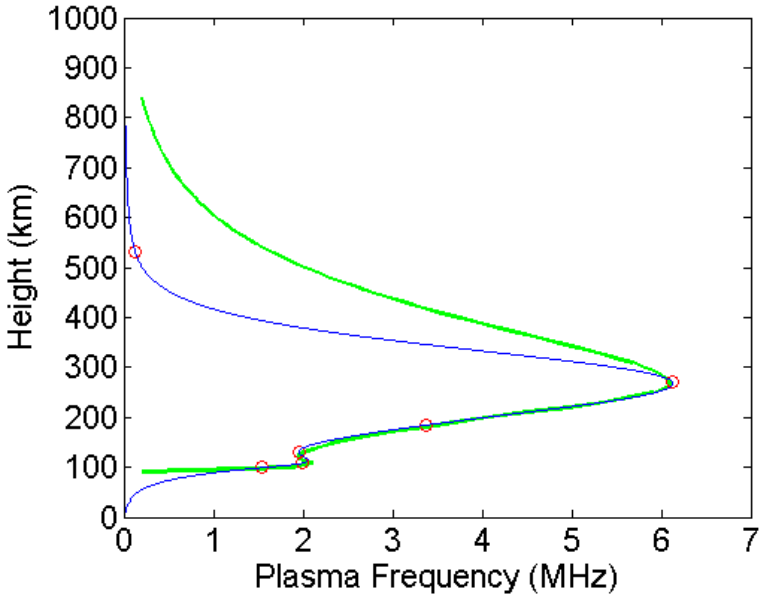

Figure 4.12: Final Booker profile (blue) using six nodes (red circles), and ARTIST-5's profile (green), for May 29.

ionospheric electron density profile entails heating at full power, in O-mode, at frequencies just below the F2 peak density plasma frequency, $foF2$. Higher rates of success for creating FAI's have been observed for sounding frequencies

that also sit near an electron gyroharmonic for the target altitude.

4.9.1 Conjunction Simulation

The experiment design required to estimate HAARP-induced irregularities based on GPS data is illustrated in Fig. 4.14. The principle requirement for the experiment is to have the perturbed zone of the ionosphere be traversed by one or more LOS paths between GPS satellites and available ground-based receivers. Unless these receivers are capable of tracking P(Y) code, the satellites must transmit dual-frequency civilian GPS signals, the L1 C/A code signal and the new L2C signal, the latter being available only on a subset of the current GPS constellation. Fig. 4.14 shows ionospheric pierce points (IPPs), where the LOS between a receiver-satellite pair crosses an ionospheric height h_{iono}. The two IPPs in Fig. 4.14 are shown as '×' symbols. The value h_{iono} for the rest of this chapter may be taken to be somewhere between 200 km and 400 km, in the F-region.

Conjunction simulations have been developed as an aid to determining the time frames and angles for aiming the HF pump, in order to perform HAARP heating experiments during times of GPS LOS conjunction with the disturbed zone of the ionosphere, taking into account a number of receivers in the vicinity of the heater array. Part of the conjunction calculation uses a Newton-Raphson algorithm to solve

$$\boldsymbol{r}_{IPP} = \boldsymbol{r}_{RX} + d\hat{\boldsymbol{r}}; \quad \hat{\boldsymbol{r}} = \frac{\boldsymbol{r}_{sat} - \boldsymbol{r}_{RX}}{||\boldsymbol{r}_{sat} - \boldsymbol{r}_{RX}||} \tag{4.32}$$

for d, which is the distance along the LOS direction $\hat{\boldsymbol{r}}$ between the receiver location r_{RX} and the GPS satellite location r_{sat} that reaches the ionospheric height

h_{iono}. The initial guess of d in the Newton-Raphson approach is that for a spherical Earth, found by solving a quadratic equation:

$$d^2 + 2\hat{\boldsymbol{r}}^T \boldsymbol{r}_{RX} d - (2\|\boldsymbol{r}_{RX}\| h_{iono} + h_{iono}^2) = 0 \tag{4.33}$$

The Newton method finds the actual IPP location above a WGS-84 ellipsoid by minimizing the difference between the WGS-84 altitude and desired altitude, $h_{wgs84}\left[\boldsymbol{r}_{IPP}(d)\right] - h_{iono}$.

When the optimal $\boldsymbol{r}_{IPP}$ has been computed for each satellite that is in view and within the 30° off-zenith range of the heater array at h_{iono}, the best heat point is found using an ellipsoid to represent the FAI, the centre of which represents the heat point. The ellipsoid can be expressed in quadratic form:

$$E = 1 - \left\| \begin{bmatrix} 1/L_X & 0 & 0 \\ 0 & 1/L_Y & 0 \\ 0 & 0 & 1/L_Z \end{bmatrix} U(\hat{\boldsymbol{b}})(\boldsymbol{r} - \boldsymbol{r}_c) \right\|^2 \tag{4.34}$$

where the point $\boldsymbol{r}$ is within the ellipsoid if $E \geq 0$. L_X, L_Y, and L_Z are the lengths of the ellipsoid's semi-axes semi-axes (i.e., major, intermediate, and minor). The radius of the ellipsoid along each of its short horizontal axes, L_Y and L_Z, is set at 15 km, such that the 30 km diameter is approximately the half-width of the HAARP heater beam. The length of the ellipsoid along the magnetic field axis, L_X, is set somewhat arbitrarily at 30 km for an elongated shape. The 3-by-3 direction cosines matrix $U(\hat{\boldsymbol{b}})$ is a rotation that transforms a position vector relative to the ellipsoid centre, the vector $\boldsymbol{r} - \boldsymbol{r}_c$, into coordinates aligned with the local magnetic field, which points in the $\hat{\boldsymbol{b}}$ direction.

The procedure for designating a heating point, or the centre of this ellipsoid, is as follows. First, if there is only one IPP in the 30° circle, $\boldsymbol{r}_c$ is trivially set

at $\boldsymbol{r}_{IPP}$. Note that in this case, it is best to choose an IPP that crosses closest to the heater's magnetic zenith (14° off-zenith, 202° azimuth for HAARP), along which the heater beam has the most effect. This is done by simply inspecting the straight-line distance between the IPP and the IPP of magnetic zenith. Fig. 4.13 is a visual representation of these distances over a 24-hour period. When more than one IPP is present, the simulation minimizes the quadratic term in Eqn. (4.34) over several IPPs $\boldsymbol{r} = \boldsymbol{r}_{IPP,1}, ..., \boldsymbol{r}_{IPP,n_E}$, where n_E becomes the maximum number of IPPs that are close enough to fit within the ellipsoid. The minimizing $\boldsymbol{r}_c$, or in other words, the point that is optimally close to the largest cluster of IPPs, is the heat point.

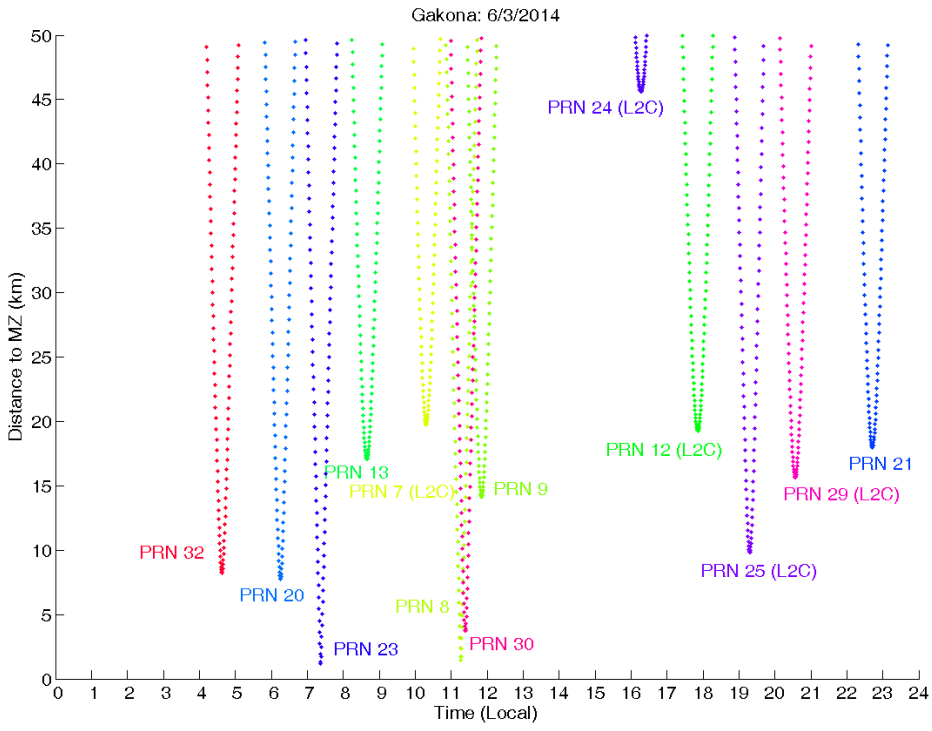

Figure 4.13: Example plot of distance to magnetic zenith for GPS satellite IPPs at 220 km, over the course of a day. Each GPS satellite is shown in a different colour.

The output of the simulation is shown in Figs. 4.14 and 4.15, where the

ellipsoid "bubble", depicted with its X semi-axis aligned with the brown geomagnetic field line, envelopes two IPPs: One along the LOS from a receiver in Yukon to PRN 24, drawn as the green dotted line, and another along the LOS from a receiver by the HAARP facility in Gakona to PRN 12, drawn as the magenta dotted line. Note that there is also an IPP associated with Fairbanks (see red "X" within the 30° off-zenith range as denoted by the magenta circle), but the conjunction with the most IPPs has been favoured by the simulation. Other IPPs, corresponding to the various receivers in view (coloured dots), are present outside of the 30° circle, and are therefore ignored and not shown. Figure 4.15 is a close-up of the heating zone and chosen conjunction.

4.9.2 Field-Aligned Irregularity Model

The electron density model for an FAI derives from a "waterbag" model [15], which prescribes a sharp jump at the border of the enhancement/depletion region. The region is modeled as being ellipsoidal in shape, with the long axis of the ellipsoid aligned with the Earth's magnetic field. The total electron density along the GPS signal path s can then be written as

$$N(\boldsymbol{r}, \boldsymbol{p}^*, \boldsymbol{x}_{FAI}) = N_0(\boldsymbol{r}, \boldsymbol{p}^*)\,[1 + (\Delta N) S\,(E(\boldsymbol{r}))] \qquad ; \quad \Delta N \geq -1 \tag{4.35}$$

$$S\,(E) = \left[1 + e^{-2\alpha_S E}\right]^{-1} \tag{4.36}$$

where ellipsoid function E comes from Eqn. 4.34, and $N_0(\boldsymbol{r}, \boldsymbol{p}^*)$ is the quiescent electron density calculated using the SREKF from the previous section. If $\boldsymbol{p}_i$ and $\boldsymbol{p}_f$ are, respectively, the background profile parameter vectors immediately prior to and following the heating experiment, then $\boldsymbol{p}^*$ is interpolated between $\boldsymbol{p}_i$ and $\boldsymbol{p}_f$ over time. $\boldsymbol{x}_{FAI} = [\boldsymbol{r}_c \quad \boldsymbol{L} \quad \Delta N]^T$ is the new state vector for the FAI

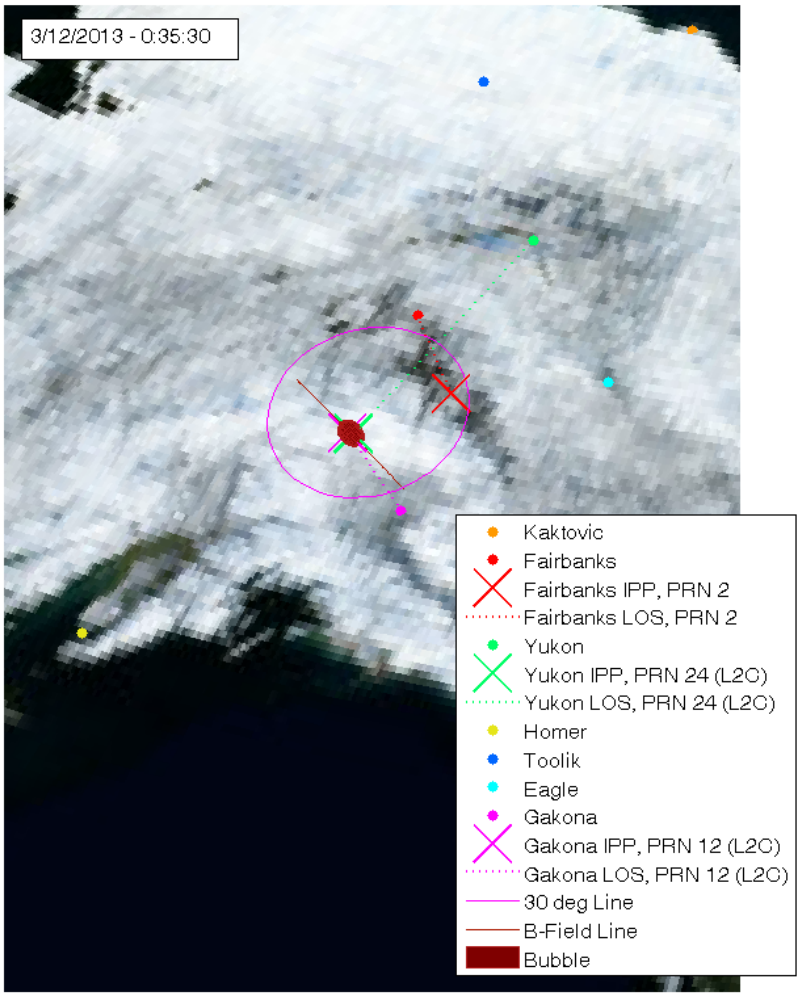

Figure 4.14: Simulation of L2C-capable GPS ray path IPPs near the heater array. PRN 12 and PRN 24 are crossing a HAARP-induced electron density irregularity bubble to receivers at Gakona and Fort Yukon, respectively, at the given time.

estimation problem. It consists of the coordinates of the ellipsoid centre, $\boldsymbol{r}_c$, a vector that consists of the 3 ellipsoid semi-major axes $\boldsymbol{L} = [L_X \quad L_Y \quad L_Z]^T$, and the fractional change in electron density within the perturbed region, ΔN. This last quantity is bound in Eqn. (4.35), as density cannot be less than zero.

For smooth integration, Eq. (4.36) is a smooth numerical approximation of the Heaviside function, with sharpness factor α_S. It transitions from the region

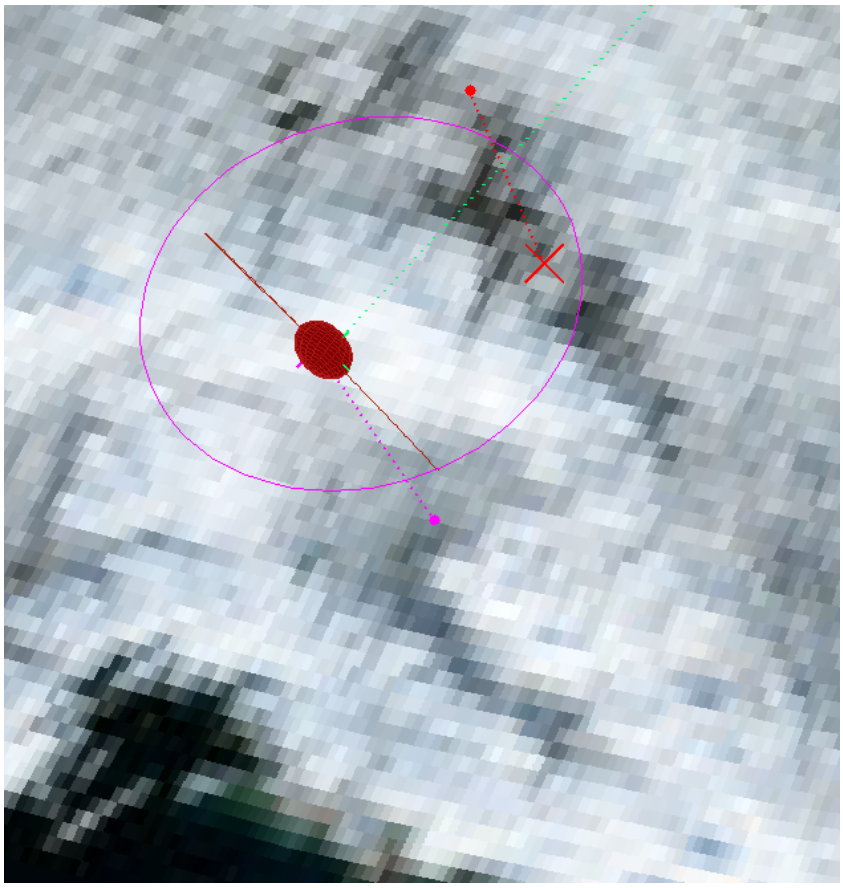

Figure 4.15: Zoomed-in view of Fig. 4.14.

outside the "waterbag", $E < 0$, to the region inside, $E \geq 0$, smoothly, as shown in the example $S(E)$ vs. E functions of Fig. 4.16 for various sharpness factors.

The new measurement model is then

$$\boldsymbol{y}_{FAI} = \boldsymbol{h}(\boldsymbol{x}_{FAI}) + \boldsymbol{w}_{FAI} = \begin{bmatrix} \Delta \boldsymbol{P}'^{1}_{fGPS}(\boldsymbol{x}_{FAI}) \\ \vdots \\ \Delta \boldsymbol{P}'^{n_{GPS}}_{fGPS}(\boldsymbol{x}_{FAI}) \end{bmatrix} + \boldsymbol{w}_{FAI} \tag{4.37}$$

where the stacked measurements in the vector $\boldsymbol{h}(\boldsymbol{x}_{FAI})$ (and their sensitivities) are defined by Eq. (4.25), except the unknown state variables on which the carrier-smoothed differential group delays depend are now contained in the perturbation state of the FAI irregularity, $\boldsymbol{x}_{FAI}$. As with the GPS differential-

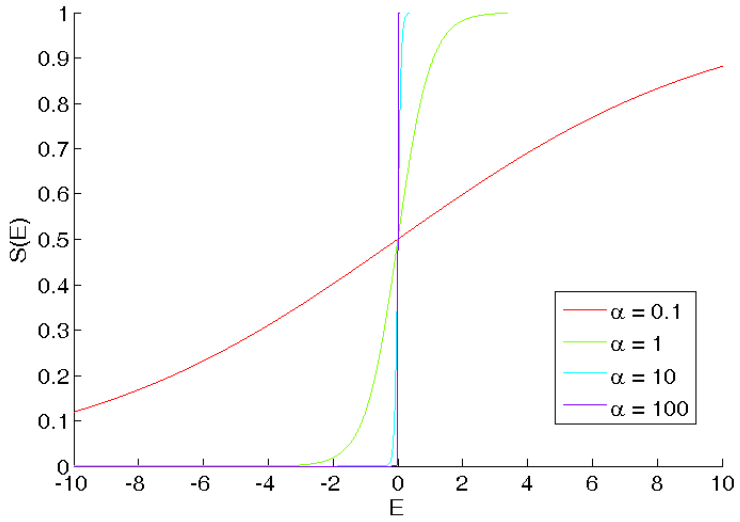

Figure 4.16: The function $S(E)$, given in Eq. (4.34), which imparts the "waterbag" behaviour in Eq. (4.35), with a few values of the sharpness factor.

frequency observables model for the quiescent ionosphere, the dependence of these quantities on the elements of $\boldsymbol{x}_{FAI}$ is based on the dependence of the electron density profile on these elements and the effect of the electron density profile on the integral in Eq. (4.23).

4.10 Ionospheric Irregularity Parameterization and Irregularity Estimation using GPS Signals, Results

Multiple signal conjunctions crossing the FAI increases the observability of the parameters in $\boldsymbol{x}_{FAI}$. However, due to the low availability of data collection campaign times that coincide with L2C satellite passes over such high latitudes, as well as limited receiver locations, the following analysis of experimental data will consider cases with only a single satellite conjunction. Data from other receivers nearby were also collected, but showed no sign of crossing the FAI. A

higher number of viable conjunctions might require more strategic placement of receivers.

Similar to the quiescent estimation results of Section 4.8, the experiments were conducted on April 29, May 29, May 30, and June 9, 2014, UTC. Variations in the data from these particular cases suggest the creation of FAIs, which were further confirmed by other instruments at HAARP (HF radio soundings/backscatter, ultra-high frequency radar backscatter, and optical imaging). The FAIs, which all happen to be enhancement regions, were formed by heating at a constant frequency around 5.6 MHz, corresponding to the fourth electron gyroharmonic at peak F2 height, which, for these cases, coincided with peak plasma frequencies (*foF2*'s) of approximately the same magnitude. Heating was started at least 10 minutes prior to the predicted conjunction, allowing FAI formation to complete. Figures 4.17 and 4.18 are snapshot examples of their LOS and IPP configurations, rendered in the same manner as in Section 4.9.1, except that the ellipsoids have been drawn according to their estimated parameters. These two figures apply to the data from April 29 and May 29. The May 30 and June 9 geometries are similar to that shown in Fig. 4.18.

The batch estimation problem in Eq. (4.37) for $\boldsymbol{x}_{FAI}$ has been solved using the trust-region-reflective algorithm in MATLAB's optimization toolbox. This algorithm is similar to the Levenberg-Marquardt algorithm, except instead of updating the Levenberg-Marquardt parameter at each iteration, a direct bound on the search step is updated. An early survey is given in Ref. [29]. The MATLAB version is based on a 2-dimensional subspace approach [8], and is a simple but robust solution for this application.

Due to the limited number of GPS lines of sight through the disturbed re-

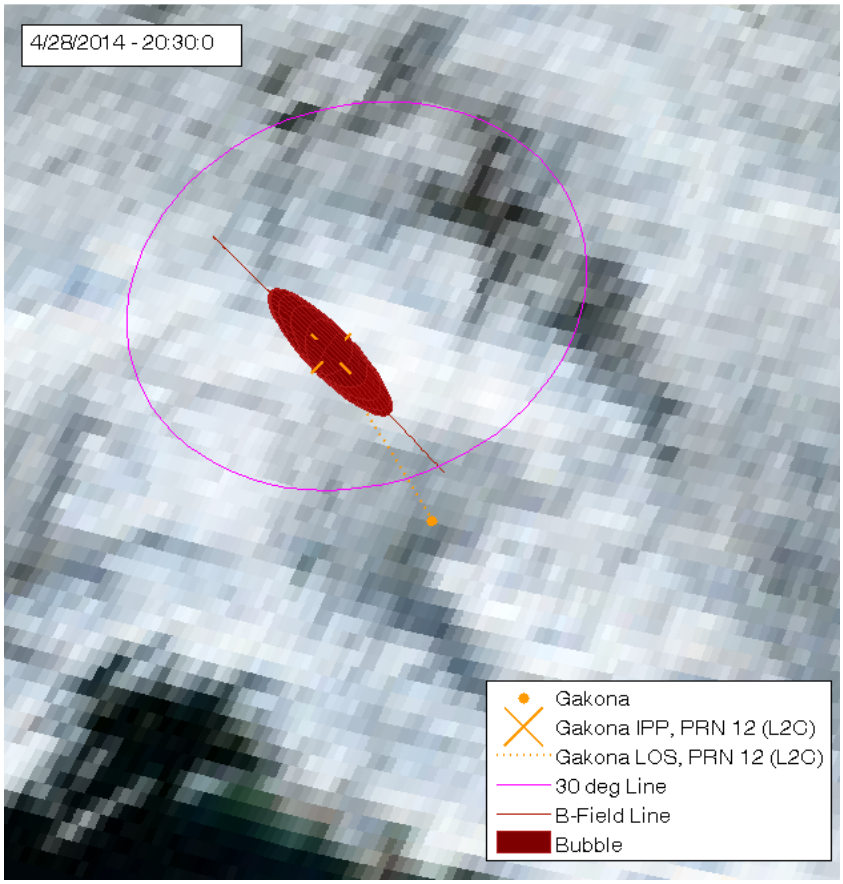

Figure 4.17: Geometry (ellipsoid width adjusted from results) of PRN 12 signal crossing the electron density irregularity bubble to a receiver at Gakona on April 29.

gion, for the sake of observability, $\boldsymbol{x}_{FAI}$ has been truncated to include only the perturbation fraction ΔN and FAI width $W = L_Y = L_Z$, while $\boldsymbol{r}_c$ is taken to be the approximate centre of the heated zone, and L_X is also a fixed guess. Assuming a mostly stable plasma cloud and slow recombination, once the FAI is formed, electron density within the cloud should be roughly constant [3], and therefore ΔN is not time-varying. The effects of drift, convection, etc. on the other parameters have been ignored for simplicity, but may be taken into account in future work, with more elaborate models, and observed through more conjunctions.

Figures 4.19, 4.20, 4.21, 4.22, and 4.23 are plots of the TEC fittings for the four

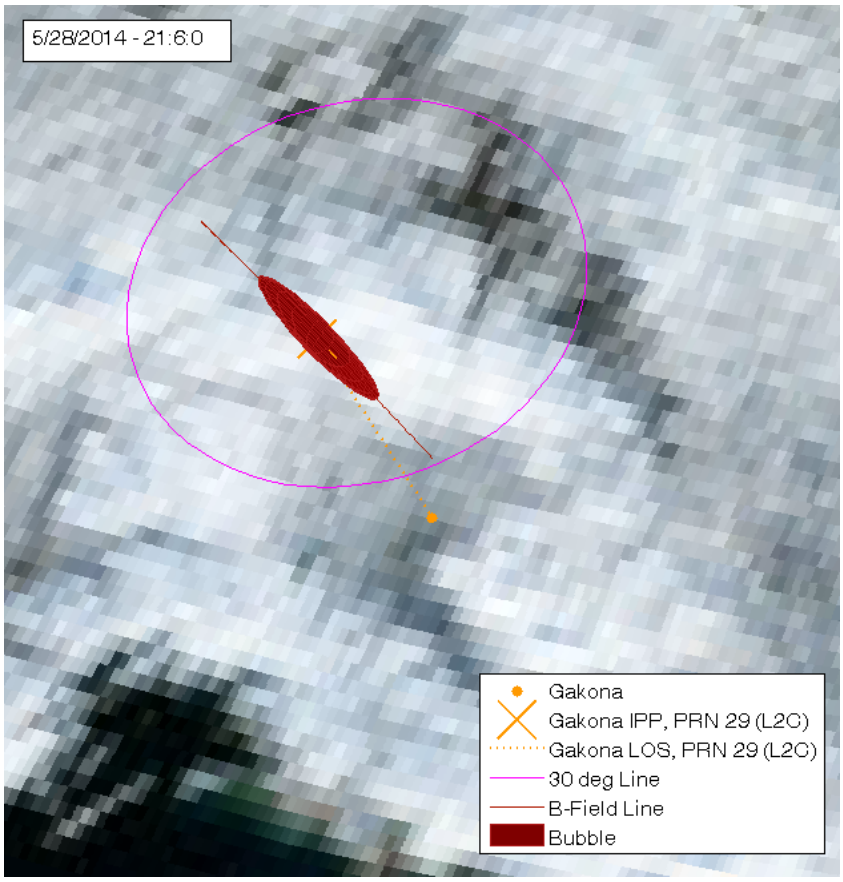

Figure 4.18: Geometry (ellipsoid width adjusted from results) of PRN 29 signal crossing the electron density irregularity bubble to a receiver at Gakona on May 29.

experiments that resulted in FAIs. The dotted vertical lines are the start and end times of the experiment.

FAI parameters $W = 25.84\ km$ and $\Delta N = 0.20$ yield the TEC fitting for PRN 12 in Fig. 4.19, using the long ellipsoid semi-axis radius $L_X = 50\ km$. At this IPP elevation, shorter lengths produce more rounded TEC offsets, while longer lengths yield flatter offsets, as seen in Fig. 4.20, which plots the same case with $L_X = 500\ km$, resulting in $W = 26.06\ km$ and $\Delta N = 0.09$. Both fits are close, with the former having an RMS fit error of 0.053 $TECU$, and the latter having a slightly higher RMS error of 0.059 $TECU$. The poorer fits after the second

dotted line, along with the rapid drop-off in TEC perturbation, suggests that the heating experiment ended before the LOS of PRN 12 completely passed through the FAI.

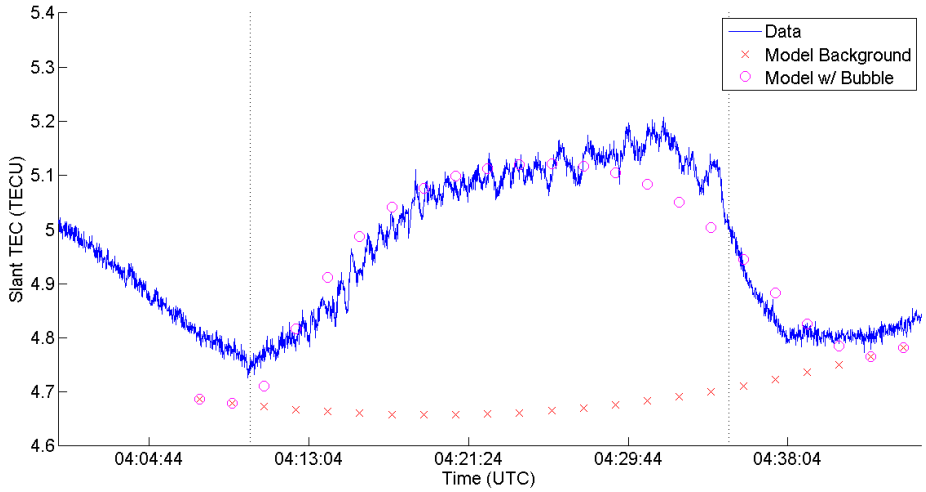

Figure 4.19: FAI model fit of TEC data for PRN 12, collected at Gakona on April 29, 2014, for a shorter semi-axis radius of 50 km. Data is compared with fit based purely on an interpolation of the quiescent ionosphere fits based on data before or after the HAARP heating event.

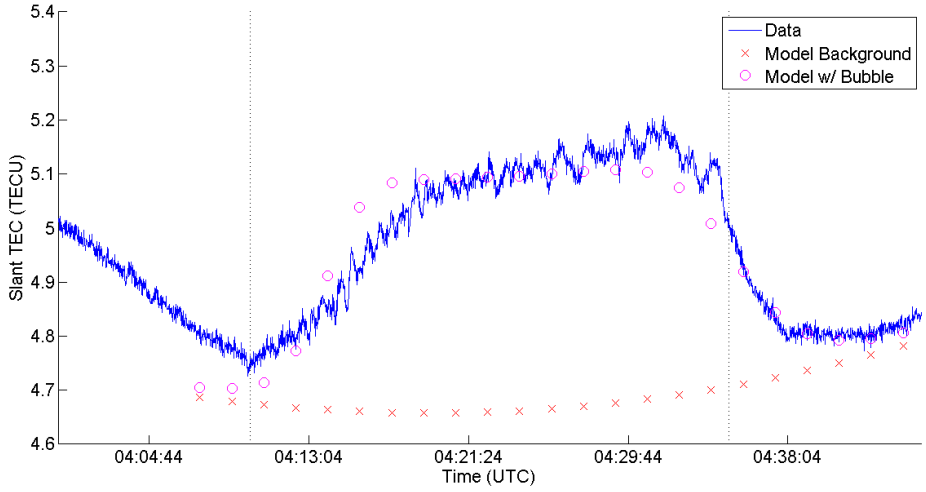

Figure 4.20: FAI TEC fit for the same data as in 4.19, except with longer semi-axis radius of 500 km, which produces a flatter TEC perturbation at this GPS satellite elevation.

Figure 4.21 shows the TEC fittings for PRN 29, for optimized values of $W = 12.99\ km$ and $\Delta N = 0.14$, using $L_X = 50\ km$. The RMS fit error for these values is 0.081 $TECU$. The dip in TEC occurring promptly after heating may be a relaxation effect. It indicates a short-coming of this method's assumption that ΔN is a constant during the event.

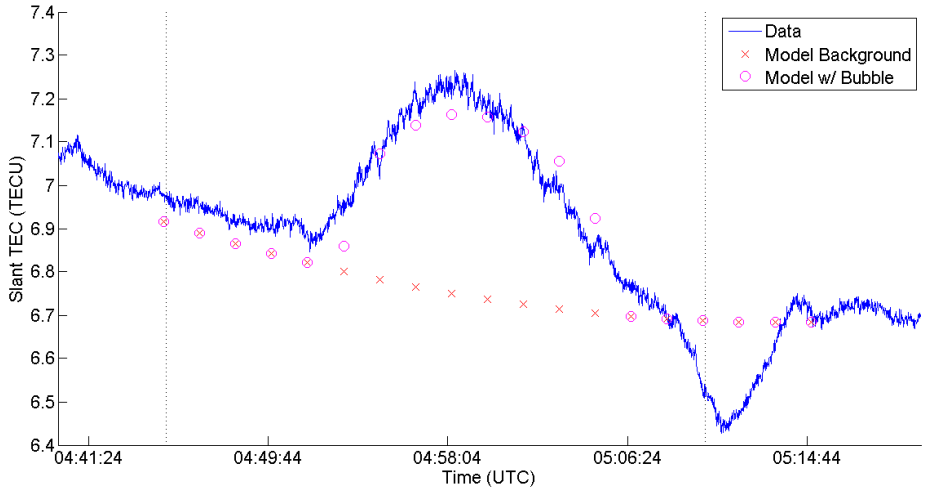

Figure 4.21: TEC fits for FAI created on May 29, 2014, at Gakona, for PRN 29.

The TEC fittings for PRN 29 in the third case can be seen in Fig. 4.22. The FAI parameters were determined to be $W = 12.99\ km$ and $\Delta N = 0.14$, for $L_X = 50\ km$. This case is an example where the start and end background TEC fits were reasonable, but not congruent enough with the data to obtain a precise perturbation fit, resulting in a RMS fit error of 0.24 $TECU$. The last case is shown in Fig. 4.23, where the nonlinear least-squares algorithm yielded $W = 18.32\ km$ and $\Delta N = 0.05$, for $L_X = 50\ km$. The RMS fit error is 0.020 $TECU$.

The TEC fits in this section are largely a proof-of-concept. Additional satellite LOS's going through other regions of the FAI would provide more information about the truncated states. Moreover, the quiescent ionospheric model

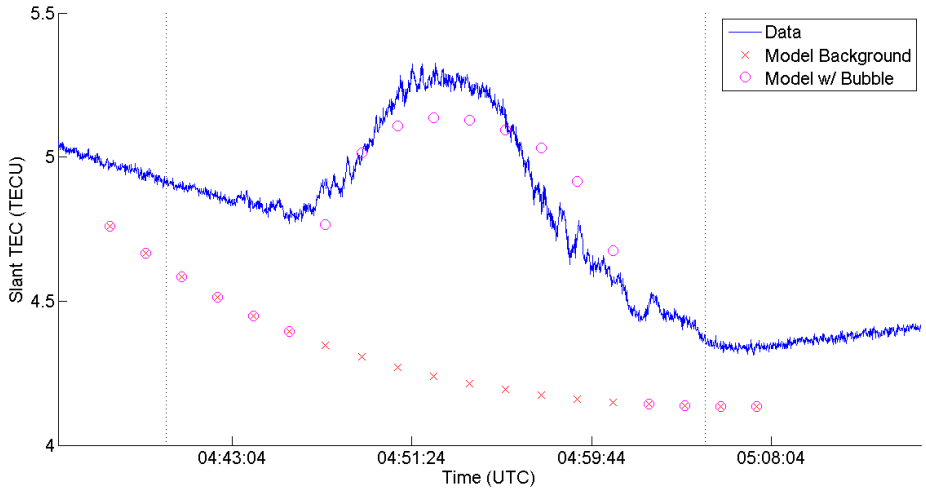

Figure 4.22: TEC fits for FAI created on May 30, 2014, at Gakona, for PRN 29.

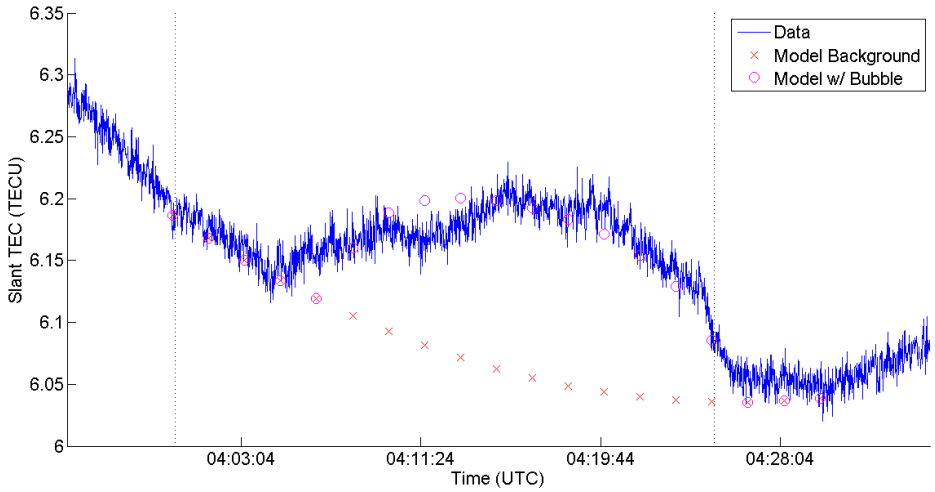

Figure 4.23: TEC fits FAI created on June 9, 2014, at Gakona, for PRN 29.

does not seem to explain everything that contributes to the TEC measurements during the quiescent time. Attempting to fit this model to many satellites means that the TEC prediction at any given point may be quite incongruous with the actual measurement. For the FAI conjunction, this translates into discrepancies before and after heating events, which can be seen distinctly in Figs. 4.19 - 4.22. Consequently, part of the optimized ΔN estimate may be compensating for

these differences. Even if the quiescent fits were improved, perhaps by better handling of the nonlinearities or additional profile parameterization, it may still be that the raw slant TEC measurements are not sufficient for the FAI problem. A suggestion for future work is to instead estimate the differences between the quiescent and disturbed slant TEC measurement time histories, and use these differences to drive the estimation of x_{FAI}. This way, the optimized parameters in x_{FAI} themselves would not be affected by the bias. Otherwise, the ellipsoidal "waterbag" model seems to describe the overall TEC disturbance well for one conjunction.

4.11 Conclusion

Refractive ray-tracing simulations of ionosonde ray paths that implement Hamilton's equations are combined with GPS phase and pseudorange perturbation integrations in a predictive measurement model. The model is used as the measurement model within a fairly standard optimal estimation/Kalman-filtering framework to fuse data from these two sources in order to estimate the parameters that characterize a quiescent electron-density profile in a nonlinear Square Root Information Filter. A test case shows that the filter produces good fits of ionosonde virtual height, to within a few kilometres. Slant total electron content (TEC) for four satellites have also been fit to within a few thousandths of TEC units.

The resulting Booker profile of the quiescent ionosphere serves as the background density function for the estimation of perturbational parameters of ionospheric field-aligned irregularities, which are characterized using an ellipsoidal

waterbag model. Dual-frequency carrier-phase-smoothed pseudorange data have been collected for a GPS signal that traverses such an irregularity, created by controlled heating with an HF pump. STEC fits characterize the effectiveness of the irregularity model and of the associated parameter estimation algorithm. Results have been generated using only a single GPS LOS conjunction with the region of disturbed ionosphere. The efficacy of this method is expected to improve the with an increased number of signal conjunctions in future experiments, and perhaps also from higher complexity in the FAI model.

CHAPTER 5

SUMMARY, CONCLUSIONS, AND FUTURE DIRECTIONS

This dissertation deals with 3 estimation systems that rely on GPS data to infer properties of signals, vehicles, or the environment. The first project involved a combined phase-locked-loop/delay-locked-loop (PLL/DLL) algorithm that tracks outputs from a bank of Doppler/code-delay correlators, even with large uncertainties in this space. It is envisioned to be useful for limb-scanning applications of GNSS signals, as they experience large degrees of path bending in these scenarios, leading to large perturbations in carrier phase and code phase that are difficult for a traditional receiver to track and that are especially difficult to track accurately very soon after initial signal acquisition. The heart of the proposed solution to this challenge is to span these uncertainties using a correlator bank, achieving considerable pull-in ranges with a hybrid acquisition/tracking algorithm. This algorithm can be quite economical for recovering meteorological information at the first rise of the GNSS satellite, but the bottleneck is processing speed. Going forward, the PLL/DLL should be evaluated using real limb scan data and an efficient, parallel Levenberg-Marquardt algorithm implementation, perhaps implemented on a graphics processing unit.

The second project used an adaptation of the first in order to solve a difficult sounding rocket attitude determination problem that involved large data gaps, many additional integer ambiguities in the beat carrier phase data due to these gaps, and under-sensing of the 3-axis attitude due to using only a 2-antenna system with the antennas displaced along a single baseline in rocket body coordinates. The three-dimensional orientation and spin rate of the sounding rocket were estimated from carrier phase differential GPS information using

a nonlinear batch filter with mixed real-valued unknowns and integer-valued unknowns. The filter's successful operation has been verified by comparing its 3-axis attitude estimates with independent estimates that were generated by a different attitude determination filter that relied on magnetometer and horizon crossing indicator data [20]. As with the first project, if this algorithm is to be used in real-time, it might be advisable to implement it using parallel processing techniques in order to enable sufficiently fast execution. As it stands, the attitude results serve well to support other remote sensing experiments on spin-stabilized vehicles like the MICA sounding rocket experiment.

The last project employed a measurement model that computes Hamiltonian-based ionosonde ray paths and their group delays, along with integrated GPS differential-frequency slant TEC observables for a given ionospheric parameterization. The optimal ionospheric profile parameterization, estimated from ionosonde and GPS data using a Square-Root Extended Kalman Filter, serves as the ambient density for later use in characterizing a field-aligned irregularity (FAI) with a simple ellipsoid "waterbag" model. A batch filter has been applied to characterize the parameters of the FAI "waterbag" model. This filter operates on slant TEC data as computed from frequency-differenced, carrier-smoothed pseudorange from a GPS satellite whose LOS to the receiver passed through the FAI. Some promising fits have been achieved using the "waterbag" ellipsoidal modal of the FAI. Better observability of all possible FAI parameters requires an expansion in visibility of GPS satellites whose LOS ray paths cross through the FAI, likely achievable by adding many more GPS receiver locations. Furthermore, other ionospheric parameterizations could also be tested in addition to the background Booker model, and models for other types of irregularities could be investigated.

APPENDIX A

CATALOG OF IONOSPHERIC TOMOGRAPHY DATA

The data collected during the HAARP campaigns currently reside on the Cornell GPS Laboratory Titanium server. Figure A.1 illustrates how the data is organized. There is a folder for each day that contains ionosonde, GPS, and other data. The ionosonde zip file contains an RSF, SAO, and PNG file for each sounding performed that day. RSF files consist of raw, multi-echo ionosonde data that can be viewed in SAO Explorer, which can be downloaded at: `http://ulcar.uml.edu/SAO-X/SAO-X.html`

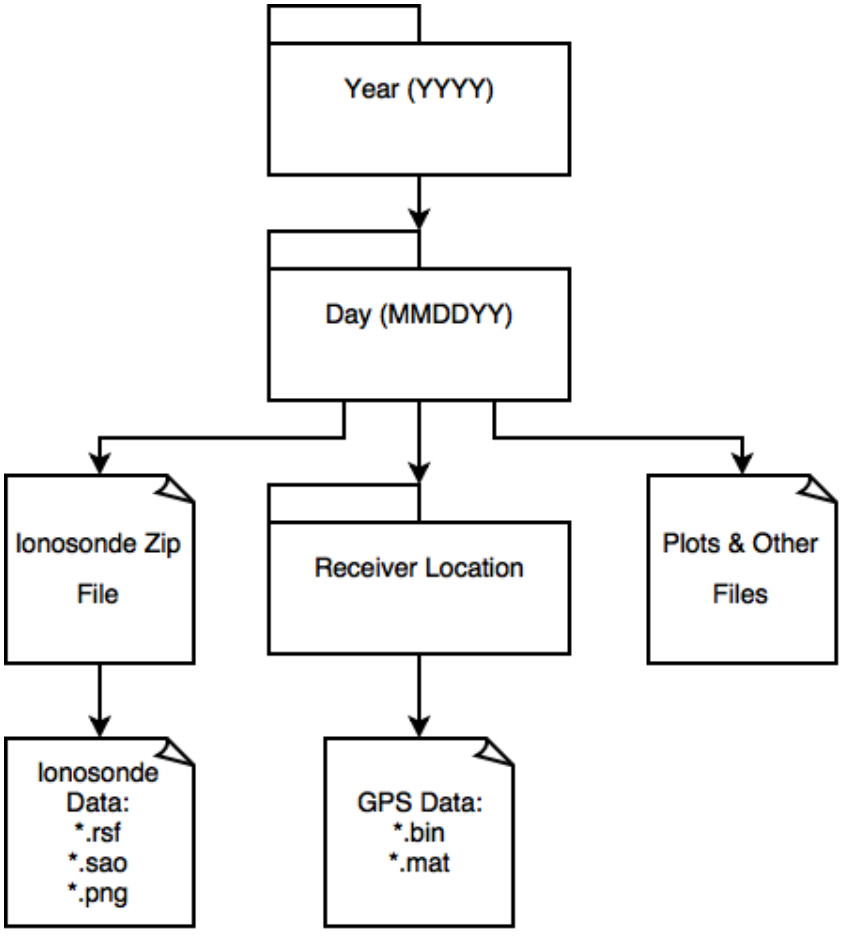

Figure A.1: Data tree for HAARP campaign data.

SAO Explorer is also capable of generating and viewing ionograms and SAO files, the contents of which are manual/software-fitted virtual height and true profile information, as well as various profile characteristics. The SAO format is described here: `http://ulcar.uml.edu/~iag/SAO-4.htm`. To analyze soundings of interest, it might be wise to first reprocess the SAO file with SAO Explorer in order to manually tweak virtual height data that has been produced by a bad fit with ARTIST-5. To load SAO information into MATLAB, there is also a ”readsao.m” function in the ”Toolbox” folder.

GPS data are separated into folders by receiver location. Depending on the receiver, the data may be in binary BIN or MATLAB MAT format. If it is a BIN file, the data must first be extracted via *binflate* in the Titanium server's standard bin directory.

Other files that may also be of use are also available for some days. These include pre-processed plots, GPS navigation files, and GPS Ionex files.

BIBLIOGRAPHY

[1] P. Axelrad and C. P. Behre, "Attitude estimation algorithms for spinning satellites using GPS phase data," *Journal of Guidance, Control, and Dynamics*, vol. 20, no. 1, pp. 164–169, Jan.-Feb. 1997.

[2] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York: John Wiley and Sons, 2001, pp. 113, 371–395.

[3] P. A. Bernhardt, "Artificially created negative-ion clouds," *Journal of Geophysical Research*, vol. 93, no. A8, pp. 8696–8704, 1983.

[4] M. Z. H. Bhuiyan and E. S. Lohan, "Advanced multipath mitigation techniques for satellite-based positioning applications," *International Journal of Navigation and Observation*, no. 412393, 2010.

[5] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*. New York: Dover Publications, 1977, pp. 69–112, 155–122, 214–217.

[6] H. G. Booker, "Fitting of multi-region ionospheric profiles of electron density by a single analytic function of height," *J. Atmos. Terr. Phys.*, vol. 39, pp. 619–624, 1977.

[7] R. G. Brown and P. Y. C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*. New York: J. Wiley, 1992, pp. 428–432.

[8] R. H. Byrd, R. B. Schnabel, and G. A. Shultz, "Approximate solution of the trust region problem by minimization over two-dimensional subspaces," *Mathematical Programming*, vol. 40, no. 1-3, pp. 247–263, 1988.

[9] R. Chaggara, C. Macabiau, and E. Chatre, "Using GPS multicorrelator receivers for multipath parameters estimation," in *Proc. ION GPS 2002*, Portland, OR, Sept. 24-27, 2002, pp. 477–486.

[10] J. H. Chapman and E. S. Warren, "Topside sounding of the Earth's ionosphere," *Space Science Reviews*, vol. 8, no. 5-6, pp. 846–865, 1968.

[11] K. Q. Z. Chiang and M. L. Psiaki, "GNSS signal tracking using a bank of correlators," in *Proc. ION GNSS 2010*, Portland, OR, Sept. 21-24, 2010, pp. 3258–3267.

[12] ——, "Kalman filter tracking of limb scan signals using a bank of correlators," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 49, no. 1, pp. 118–133, Jan. 2013.

[13] K. Q. Z. Chiang, M. L. Psiaki, S. P. Powell, R. J. Miceli, and B. W. O'Hanlon, "GPS-based attitude determination for a spinning rocket," in *Proc. ION GNSS 2012*, Nashville, TN, Sept. 17-21, 2012, pp. 2342–2350.

[14] D. M. A. W. G. S. . D. Committee and R. Smith, *Department of Defense World Geodetic System 1984: Its Definition and Relationships with Local Geodetic Systems*, ser. DMA Technical Report. Defense Mapping Agency, 1987.

[15] J. F. Drake and J. D. Huba, "Dynamics of three-dimensional ionospheric plasma clouds," *Physical Review Letters*, vol. 58, no. 3, pp. 273–281, 1987.

[16] P. C. Fenton, Kunysz, Waldemar, Garbe, and Greg, "Using GPS for position & attitude determination of the Canadian Space Agency's active rocket mission," in *Proc. ION GPS 1998*, Nashville, TN, 1998, pp. 1791–1800.

[17] P. E. Gill, W. Murray, and M. H. Wright, *Practical Optimization*. London: Academic Press, 1981, pp. 136–137.

[18] R. R. Hatch and J. E. Knight, "Method and apparatus for smoothing code measurements in a global positioning system receiver," *U.S. Patent No. 5471217*, Nov. 28, 1995.

[19] M. M. Hoque and J. Norbert, *Ionospheric Propagation Effects on GNSS Signals and New Correction Approaches*. InTech, 2012, pp. 381–404.

[20] T. E. Humphreys, M. L. Psiaki, E. M. Klatt, S. P. Powell, and P. M. Kintner, "Magnetometer-based attitude and rate estimation for spacecraft with wire booms," *Journal of Guidance, Control, and Dynamics*, vol. 28, no. 4, pp. 584–597, July - Aug. 2005.

[21] P. Y. Hwang, G. A. McGraw, and J. R. Bader, "Enhanced differential GPS carrier-smoothed code processing using dual-frequency measurements," *Navigation*, vol. 46, no. 2, pp. 127–137, 1999.

[22] G.-I. Jee, H. S. Kim, Y. J. Lee, and C. G. Park, "A GPS C/A code tracking loop based on extended kalman filter with multipath mitigation," in *Proc. ION GPS 2002*, Portland, OR, Sept. 24-27, 2002, pp. 446–451.

[23] S. Jin, E. Cardellach, and F. Xie, *GNSS Remote Sensing: Theory, Methods and Applications*. Dordrecht: Springer, 2014.

[24] R. M. Jones and J. J. Stephenson, "A versatile three-dimensional ray tracing computer program for radio waves in the ionosphere," U.S. Dept. of Commerce, OT Report, Tech. Rep., Oct. 1975.

[25] C. H. Liu and K. C. Yeh, *Theory of Ionospheric Waves*, 1st ed. New York: Academic Press, 1972, pp. 187, 234–240.

[26] M. Martin-Neira and R. Lucas, "GPS attitude determination of spin stabilized satellites," in *Proc. ION GPS 1992*, Albuquerque, NM, 1992, pp. 757–765.

[27] P. Misra and P. Enge, *Global Positioning System: Signals, Measurements and Performance*. Lincoln: Ganga-Jamuna Press, 2006, pp. 431–492.

[28] R. H. Mitch, M. L. Psiaki, and D. M. Tong, "Local ionosphere model estimation from dual-frequency global navigation satellite system observables," *Radio Sci.*, vol. 48, pp. 671–684, 2013.

[29] J. J. Moré and D. C. Sorensen, "Computing a trust region step," *SIAM Journal on Scientific and Statistical Computing*, vol. 4, no. 3, pp. 553–572, 1983.

[30] U. of Bern, "Code – center for orbit determination in europe," http://www.aiub.unibe.ch/content/research/sataellite_geodesy/code___research/index_eng.html.

[31] B. W. O'Hanlon, M. L. Psiaki, S. P. Powell, J. A. Bhatti, T. E. Humphreys, G. Crowley, and G. S. Bust, "CASES: A smart, compact GPS software receiver for space weather monitoring," in *Proc. ION GNSS 2011*, Portland, OR, Sept. 20-23, 2011, pp. 2745–2753.

[32] T. Pany, B. Eissfeller, and G. Hein, "A two dimensional (delay/doppler) multi correlator in a multi-frequency PC-based software receiver," in *Proc. NAVITEC 2006*, Noordwijk, The Netherlands, 2006.

[33] M. L. Psiaki, "Block acquisition of weak GPS signals in a software receiver," in *Proc. ION GPS 2001*, Salt Lake City, UT, Sept. 11-14, 2001, pp. 2838–2850.

[34] ——, "Batch algorithm for global-positioning-system attitude determina-

tion and integer ambiguity resolution," *Journal of Guidance, Control, and Dynamics*, vol. 29, no. 5, pp. 1070–1079, Sept.-Oct. 2006.

[35] M. L. Psiaki and J. C. Hinks, "Numerical solution of a generalized wahba problem for a spinning spacecraft," *Journal of Guidance, Control, and Dynamics*, vol. 35, no. 3, pp. 764–773, May-June 2012.

[36] M. L. Psiaki and H. Jung, "Extended Kalman filter methods for tracking weak GPS signals," in *Proc. ION GPS 2002*, Portland, OR, Sept. 24-27, 2002, pp. 2539–2553.

[37] M. L. Psiaki and S. Mohiuddin, "Global positioning system integer ambiguity resolution using factorized least-squares techniques," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 2, pp. 346–356, Mar.-Apr. 2007.

[38] ——, "Modeling, analysis, and simulation of GPS carrier phase for spacecraft relative navigation," *Journal of Guidance, Control, and Dynamics*, vol. 30, no. 6, pp. 1628–1639, Nov.-Dec. 2007.

[39] M. L. Psiaki, J. Theiler, J. Bloch, S. Ryan, R. W. Dill, and R. E. Warner, "ALEXIS spacecraft attitude reconstruction with thermal/flexible motions due to launch damage," *Journal of Guidance, Control, and Dynamics*, vol. 20, no. 5, pp. 1033–1041, Sept.-Oct. 1997.

[40] B. W. Reinisch and X. Huang, "Deducing topside profiles and total electron content from bottomside ionograms," *Advances in Space Research*, vol. 27, no. 1, pp. 23–30, 2001.

[41] B. W. Reinisch, X. Huang, A. Belehaki, J. Shi, M. Zhang, and R. Ilma, "Modeling the IRI topside profile using scale heights from ground-based ionosonde measurements," *Advances in Space Research*, vol. 34, no. 9, pp. 2026–2031, 2004.

[42] J. Röttger, "Ionosphere and atmosphere research with radars," *Max-Planck-Institut, Lindau, UNESCO Encyclopedia of Life Support Systems, Geophysics and Geochemistry, Paris*, 2004.

[43] S. M. Stankov, N. Jakowski, and S. Heise, "A new method for reconstruction of the vertical electron density distribution in the upper ionosphere and plasmasphere," *Journal of Geophysical Research*, vol. 108, no. A5, p. 1164, 2003.

[44] S. M. Stankov, K. Stegen, P. Muhtarov, and R. Warnant, "Local ionospheric electron density profile reconstruction in real time from simultaneous ground-based GNSS and ionosonde measurements," *Advances in Space Research*, vol. 47, no. 7, pp. 1172–1180, Apr. 2003.

[45] P. J. G. Teunissen, "The least-squares ambiguity decorrelation adjustment: A method for fast GPS integer ambiguity estimation," *Journal of Geodesy*, vol. 70, no. 1-2, pp. 65–82, Nov. 1995.

[46] B. R. Townsend, P. C. Fenton, and K. J. van Dierendonck, "L1 carrier phase multipath error reduction using MEDLL technology," in *Proc. ION GPS 1995*, Palm Springs, GA, Sept. 13-15, 1995, pp. 1539–1544.

[47] M. K. Transtrum and J. P. Sethna, "Improvements to the levenberg-marquardt algorithm for nonlinear least-squares minimization," *arXiv preprint arXiv:1201.5885*, 2012.

[48] F. Van Graas, A. Soloviev, M. U. De Haag, S. Gunawardena, and M. Braaasch, "Comparison of two approches for GNSS receiver algorithms: Batch processing and sequential processing considerations," in *Proc. ION GNSS 2005*, Long Beach, CA, Sept. 13-16, 2005, pp. 200–211.

[49] M. Wax, "The joint estimation of differential delay, doppler, and phase," *IEEE Trans. on Information Theory*, vol. IT-28, no. 5, pp. 817–820, Sept. 1982.

[50] J. R. Wertz, *Spacecraft Attitude Determination and Control*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1978, pp. 411–420.

[51] J.-H. Won, D. Dötterböck, and B. Eissfeller, "Performance comparison of different forms of Kalman filter approach for a vector-based GNSS signal tracking loop," in *Proc. ION GNSS 2009*, Savannah, GA, Sept. 22-25, 2009, pp. 3037–3048.

[52] C. Yang, "GPS signal tracking with Kalman filter based on joint code delay and carrier phase and frequency error discriminator," in *Proc. ION 60th Annual Meeting 2004*, Dayton, OH, June 7-9, 2004, pp. 631–640.

[53] J. York, J. Little, D. Munton, and K. Barrientos, "A complex-ambiguity function approach to a GPS receiver," in *Proc. ION GNSS 2009*, Savannah, GA, Sept. 22-25, 2009, pp. 2637–2645.

[54] N. I. Ziedan and J. L. Garrison, "Extended Kalman filter-based tracking

of weak GPS signals under high dynamic conditions," in *Proc. ION GNSS 2004*, Long Beach, CA, Sept. 21-24, 2004, pp. 20–31.